RE
Transactions
n ANTENNAS and PROPAGATION

Volume AP-7 OCTOBER, 1959 Number 4

Published Quarterly

TABLE OF CONTENTS

CONTRIBUTIONS

COMMUNICATIONS

PUBLISHED BY THE
ofessional Group on Antennas and Propagation

IRE TRANSACTIONS® PGAP IS A PUBLICATION DEVOTED TO EXPERIMENTAL AND THEORETICAL PAPERS ON RADIO ANTENNAS, ON GUIDED OR UNGUIDED PROPAGATION OF RADIO WAVES, AND ON ALLIED FIELDS OF RADIO PHYSICS SUCH AS RADIO ASTRONOMY

MANUSCRIPTS should be submitted to John B. Smyth, Editor, Smyth Research Associates, 3555 Aero Court, San Diego 11, Calif. Manuscripts should be original typewritten copy, double spaced, plus one carbon copy. References should appear as footnotes and include author's name, title, journal, volume, initial and final page numbers, and date. Each paper must have a summary of not more than 200 words. News items concerning PGAP members and group activities should be sent to the News Editor, Mr. Arthur Dorne, Dorne and Margolin, Inc., 30 Sylvester Street, Westbury, L.I., N.Y.

ILLUSTRATIONS should be submitted as follows: All line drawings (graphs, charts, block diagrams, cutaways, etc.) should be inked uniformly and ready for reproduction. If commercially printed grids are used in graph drawings, author should be sure printer's ink is of a color that will reproduce. All halftone illustrations (photographs, wash, airbrush, or pencil renderings, etc.) should be clean and ready to reproduce. Photographs should be glossy prints. Call-outs or labels should be marked on a registered tissue overlay, not on the illustration itself. No illustration should be larger than 8 x 10 inches.

Copies can be purchased from **THE INSTITUTE OF RADIO ENGINEERS, 1 East 79 St., New York 21, N.Y. PRICE PER COPY:** members of the Professional Group on Antennas and Propagation, $2.55; members of the IRE, $3.85; nonmembers, $7.65. **ANNUAL SUBSCRIPTION PRICE:** PGAP members, included in PGAP fee of $4.00; IRE members, $8.50; Colleges and public libraries, $10.00; nonmembers, $17.00. IRE TRANSACTIONS ON ANTENNAS AND PROPAGATION.

Entered as second-class matter, at the post office at Menasha, Wisconsin, under the act of August 24, 1912. Acceptance for mailing at a special rate of postage is provided for in the act of February 28, 1925, embodied in Paragraph 4, Section 412, P. L. & R., authorized October 26, 1927.

AN IMPORTANT ANNOUNCEMENT
TO ALL PGAP MEMBERS AND SUBSCRIBERS

The IRE Professional Group on Antennas and Propagation will shortly publish the *Proceedings of the URSI International Symposium on Electromagnetic Theory,* held at the University of Toronto, Canada, on June 15-20, 1959. It will be published as a special supplement to Volume AP-7 (1959) of the IRE TRANSACTIONS on Antennas and Propagation.

Those who registered at the Toronto Symposium will automatically receive one copy as a part of their symposium registration fee. PGAP members and TRANSACTIONS subscribers may obtain a copy by ordering at the rates indicated below. There will be no free distribution because of the special nature and large size of the supplement.

This imposing 400-page volume, comprising invited papers by 54 of the world's leading authorities, promises to be one of the outstanding reference works in its field. The subjects covered include Diffraction and Scattering Theory, Radio Telescopes, Surface Waves, Boundary Value Problems, Propagation of Waves, and Antennas. The complete program may be found on page 18A of the June, 1959 issue of the PROCEEDINGS OF THE IRE.

PGAP members and subscribers are urged to order their copies now by returning the form below to the IRE, accompanied by remittance made payable to The Institute of Radio Engineers.

.....

The Institute of Radio Engineers, Inc.
1 East 79 Street, New York 21, N.Y.

ORDER FORM FOR PGAP TORONTO SYMPOSIUM PROCEEDINGS

No. of Copies	Price per Copy		Amount Paid
	PGAP Members	$ 8.00	$
	Other IRE Members	$12.00	$
	Libraries	$12.00	$
	Nonmembers	$16.00	$
		TOTAL REMITTANCE	$

ORDERED BY (Please Print):_____

SHIPPING ADDRESS:_____

_____

contributions

Leaky-Wave Antennas I: Rectangular Waveguides*

L. O. GOLDSTONE† AND A. A. OLINER†

Summary—**A microwave network approach is employed for the description and analysis of leaky-wave antennas. This approach is based on a transverse resonance procedure which yields the complex propagation constants for the leaky waves. A perturbation technique is then applied to the resonance equation to obtain results in simple and practical form. These procedures are illustrated by application to a number of practical leaky rectangular waveguide structures. Very good agreement is obtained between the theoretical results and the measured values.**

I. Introduction

IT IS KNOWN that the radiation from a continuous longitudinal slot in a uniform lossless waveguide can frequently be characterized by a traveling wave with a complex propagation constant. This traveling wave propagates along the waveguide with a velocity greater than that of light and is attenuated as it travels, thus indicating a continuous leakage of energy. Waves possessing this characteristic behavior have been designated as *leaky waves*.

These leaky waves are not characteristic modes of the open waveguide region. The nonmodal character of such waves is evident from the fact that they increase without limit in the transverse direction. This improper behavior of the leaky waves and the distinction between them and characteristic modes are discussed in the next section. Despite this unphysical behavior, however, these nonmodal waves may nevertheless be employed for the representation of field solutions in suitably restricted portions of open waveguide regions. Moreover, since the leaky waves are solutions of the source-free field equations, their propagation constants can be obtained rigorously from a transverse resonance calculation which is formally identical to that for discrete characteristic modes.

In order to carry out such a resonance calculation, however, it is first necessary to effect a *transverse network representation* of the open structure. Two steps are required in accomplishing this. The first consists of the introduction of a complete set of characteristic modes and the second entails the solution of a discontinuity problem. Different modal sets can be employed in effecting the network representation. In this paper a set of modes differing from the usual E and H (or TM and TE) modes are derived which are particularly convenient for the transverse network representation of slotted rectangular waveguides. These modes are designated as E-type and H-type modes, and are discussed in detail in Section IV. If conventional E or H modes in the transverse direction are used in the description of leaky-wave structures, the resulting transverse network may be complicated in form, involving the

* Manuscript received by the PGAP, June 9, 1958; revised manuscript received, November 3, 1958. This investigation was performed at the Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., under Contract No. AF 19(604)-2031 with the AF Cambridge Res. Ctr., Bedford, Mass. The paper is a portion of a dissertation submitted by L. O. Goldstone in partial fulfillment of the D.E.E. degree at the Polytechnic Inst. of Brooklyn, May, 1957.

† Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y.

coupling of two or more transverse modes of different types. On the other hand, the use of E-type or H-type modes results in the simplest form of network for all uniform homogeneous leaky-wave structures since these two types of modes are never coupled in such structures. For many leaky waveguides of practical interest, the second step, that associated with the discontinuity evaluation, does not need to be actually carried out since the desired results are available in the literature. When this is the case the network representation may be obtained immediately without solving a field problem. In any case, this approach reduces the calculation of the propagation constants to the solution of a *network* problem. In general, the transverse network description of a slotted waveguide as viewed from within the guide consists of a number of transmission lines coupled by a lumped network at the plane of the slot. However, in the cases of practical interest where the field is characterized by a single leaky wave the network consists of only one transmission line, short-circuited at one end and terminated in a lumped network at the other end.

The network resonance equations, which must be solved for the propagation constants, can be conveniently treated by a perturbation technique since in the cases of interest the leaky-wave propagation constants can be regarded as perturbations on the propagation constants of modes in closed waveguides with perfectly conducting or reactive walls. The major advantage of such a perturbation technique is that it yields *simple closed-form expressions* for the propagation constants, whereas exact solutions (of these *complex* transcendental equations) require either computing machines or arduous numerical computations. An added advantage is that the functional behavior of the solution is immediately evident, thus permitting a measure of insight.

It should be emphasized that the application of microwave network concepts to the analysis of leaky-wave structures involves the decomposition of the over-all problem into two essentially independent parts. One of these is a field problem for the network parameters of the appropriate waveguide discontinuity. The other is a network resonance problem which yields the desired propagation constants directly. The chief advantages of this decomposition are as follows:

1) it is a systematic procedure which reduces the over-all problem to a simple form common to a variety of structures;
2) the network parameters of a large number of waveguide discontinuities are either known or can be obtained from the application of relatively simple "small aperture" techniques, thus eliminating the necessity of treating difficult field problems;
3) the resonance calculations are conveniently treated by perturbation procedures to yield simple closed-form expressions for the propagation constants.

The analysis of leaky wave antennas has received considerable attention in the literature.[1–12] Hansen,[1] who first suggested the use of leaky structures as antennas, was apparently also the first to recognize that the behavior could be characterized by a complex propagation constant, and presented an elementary but pioneering analysis based upon this concept. The most extensive analysis of slotted rectangular waveguides has been that presented by Rumsey.[6] In contrast to the microwave network approach, however, Rumsey treats simultaneously the two aspects of these problems referred to above. His variational procedure for determining the propagation constant actually incorporates in a concealed fashion the determination of the network parameters of the transverse discontinuity. The microwave network approach makes direct use of the available results for the transverse discontinuities, thereby saving considerable effort and reducing the complexity of the form of the results. Some workers, however, have made use of transverse network representations in connection with leaky waveguides. Rotman[5] used such a network for an approximate analysis of the channel waveguide antenna (see diagram 3 in the Appendix) but he did not employ a correct transverse resonance procedure. More recently, Honey has independently employed network representations and a transverse resonance procedure for the calculation of leaky-wave propagation constants.[10,11] However, his work is confined to two specific structures (see diagram 2 in the Appendix for one of these), and the references cited contain no theoretical discussion which would provide a basis for the extension of the network approach to a wider class of leaky structures. He has not employed perturbation techniques for the solution of the resonance equations but has obtained exact numerical solutions with the aid of a computing machine.

In this paper, the following structures will be treated as examples of the network approach to the analysis of leaky rectangular waveguides: 1) slotted rectangular

[1] W. W. Hansen, "Radiating electromagnetic waveguide," U. S. Patent No. 2,402,622.
[2] H. G. Booker, "Girders and Trenches as End-Fire Aerials," Telecommun. Res. Establ., Swanage, Eng., Rept. No. 30; 1941.
[3] A. L. Cullen, "On the channel section waveguide radiator," *Phil Mag.*, vol. 40, pp. 417–428; April, 1949.
[4] A. S. Schelkunoff and H. T. Friis, "Antenna Theory and Practice," John Wiley and Sons, Inc., New York, N. Y.; 1952.
[5] W. Rotman, "The channel guide antenna," *Proc. Nat. Electronics Conf.*, vol. 5, 1949.
[6] V. H. Rumsey, "Traveling wave slot antennas," *J. Appl. Phys.*, vol. 24, pp. 1358–1365; November, 1953.
[7] R. F. Harrington, "Propagation along a slotted cylinder," *J. Appl. Phys.*, vol. 24, pp. 1366–1371; November, 1953.
[8] J. N. Hines, V. H. Rumsey and C. H. Walter, "Traveling-wave slot antennas," PROC. IRE, vol. 41, pp. 1624–1631; November, 1953.
[9] F. J. Zucker, "The Guiding and Radiation of Surface Waves," *Proc. Symp. on Modern Advances in Microwaves Techniques*, Polytechnic Institute of Brooklyn, Brooklyn, N. Y.; November, 1954.
[10] R. C. Honey, "Horizontally Polarized Long-Slot Array," Stanford Res. Inst., Menlo Park, Calif., Tech. Rept. No. 47; August, 1954.
[11] R. C. Honey, "A Flush-Mounted Horizontally Polarized Directional Antenna," Stanford Res. Inst., Menlo Park, Calif., Tech. Rept. No. 54; January, 1956.
[12] R. S. Elliott, "Serrated waveguide—part I: theory," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-5, pp. 270–275; July, 1957.

waveguide carrying the H_{10} leaky wave; 2) slotted rectangular waveguide carrying the E_{11} leaky wave; 3) rectangular waveguide with a periodic array of circular apertures in the narrow face carrying the H_{10} leaky wave. It is understood, of course, that the foregoing leaky-wave designations correspond to the unperturbed modes that would exist in the closed guide. Since the H_{10} and E_{11} modes in closed rectangular waveguide can be regarded as the lowest H-type and E-type modes, respectively, propagating transversely, the leaky waves corresponding to these modes will hereafter be referred to as H-type and E-type leaky waves. A tabular summary of the theoretical results for these structures is presented in the Appendix. These results are compared in Section VI with measured data, either previously available or obtained in connection with this study. The consistently favorable comparison between the theoretical and measured values indicates that the perturbation procedures are indeed sufficiently accurate for practical purposes.

II. Characteristics of Leaky Waves

It was mentioned in the Introduction that while leaky waves are solutions of the source-free Maxwell equations they are not proper modes because of their singular behavior at infinity in the transverse, or cross-section, plane. Despite the decidedly unphysical nature of this aspect of the leaky-wave solutions, however, they are valid field representations in certain restricted regions within which they remain finite. It is the purpose of this section to expand on these remarks.

A leaky wave travels along its guiding structure with a complex propagation constant; its phase velocity is greater than that of light and its attenuation constant is indicative of a continuous leakage of energy along the guiding structure. The improper transverse variation of the field of a leaky wave follows directly from the relationship among the wavenumbers for rectangular regions, which is

$$k^2 = k_x^2 + k_y^2 + k_z^2, \tag{1}$$

where $k=2\pi/\lambda$ is the free space wavenumber and k_x, k_y, and k_z are the wavenumbers in x, y, and z directions respectively. If the leaky wave propagates in the (longitudinal) z direction with

$$k_z = -j\gamma_z = \beta_z - j\alpha_z, \tag{2}$$

and if k_y is real and less than k, then the wavenumber in the (transverse) x direction is necessarily of the form

$$k_x = \beta_x + j\alpha_x, \tag{3}$$

with $\beta_x>0$ and $\alpha_x>0$. If the region into which the leaky structure radiates is defined by $x>0$, the variation of the leaky wave in the x direction is

$$e^{(\alpha_x - j\beta_x)x}. \tag{4}$$

It is evident, therefore, that the leaky wave propagates in the x direction with ever *increasing* amplitude, becoming infinite at infinity in the transverse plane.

Despite this improper transverse behavior, leaky waves can provide a valid and highly convergent representation of the field within certain restricted regions of source-excited open waveguide structures. This statement may be clarified by reference to Fig. 1, which

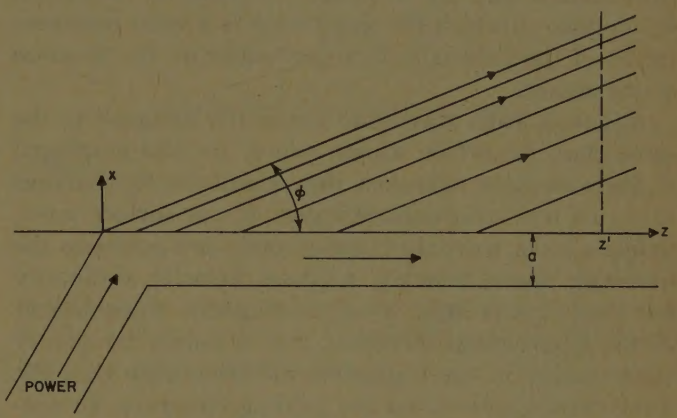


Fig. 1—Radiation from a semi-infinite leaky waveguide.

shows a semi-infinite uniform leaky waveguide of width a, fed by a closed waveguide. The radiating slot is contained in the $x=0$ plane, starts at $z=0$ and extends to infinity in the positive z direction. Since leaky waves are solutions of the homogeneous field equations the incident wave in the feed guide can excite a leaky wave within the open waveguide. In the interior of the open waveguide, $0>x>-a$, the leaky wave is finite everywhere and constitutes a valid representation of the field. This leaky wave excited within the guide radiates into the space $x>0$ in the manner indicated by Fig. 1. Any radiation from the junction between the feed guide and the leaky guide is of course neglected here. The angle at which the radiation occurs is given approximately by

$$\Phi \approx \cos^{-1}\frac{\lambda}{\lambda_g}, \tag{5}$$

where $\lambda_g=2\pi/\beta_z$ is the guide wavelength of the leaky wave in the z direction. Since the amplitude of the leaky wave decays as the wave propagates along the z direction, the amount of radiation per unit length decreases with increasing z. This variation in the intensity of radiation is indicated in Fig. 1 by the decrease in the density of the rays. It can be seen that at any point $z'>0$ the field intensity increases in the x direction up to the height $x=z'\tan\Phi$. The x dependence of the field in this region is just that given by (4). For $x>z'\tan\Phi$ the field intensity must decrease rapidly because no radiation from the leaky wave reaches this region. The radiation from the leaky wave is therefore essentially confined to the region $x<z\tan\Phi$. Everywhere in

this region the leaky wave has a finite amplitude and is therefore a valid representation of the radiation field. It can also be seen that the region in which the amplitude of the leaky wave approaches infinity ($x \to \infty$) is automatically excluded from the domain in which the radiation field can be correctly characterized by the leaky wave. One is thus confronted by the interesting feature that while the properties of the leaky wave are analytically determined by a source-independent condition, the domain in which the leaky wave is a valid representation of the field is in fact dependent on the location of the source.

Although leaky waves are frequently assigned to the same class as surface waves, which are also employed in flush-mounted antennas, they are distinctly different in nature from true surface waves. A true surface wave, unlike a leaky wave, is a proper mode and possesses the following characteristics: a phase velocity ordinarily less than that of light, a real propagation wavenumber in the longitudinal direction, and a purely imaginary wavenumber in the transverse direction such that the wave decays away from the guiding structure. In contrast, a leaky wave is nonmodal and possesses a phase velocity along the surface greater than that of light, a complex propagation constant in the longitudinal direction, and a complex transverse wavenumber such that the wave propagates transversely away from the guiding structure with increasing amplitude. In addition, these two wave types have different mechanisms of radiation. The surface wave is basically nonradiating and can produce radiation only at a discontinuity on the guiding structure whereas the leaky wave radiates continuously.

The nonmodal nature of leaky waves can be exhibited only by a careful examination of the modal spectrum of the guiding structure. The modal spectrum of an open waveguide is usually purely continuous although under certain conditions there may also exist a discrete spectrum. It can be shown[13] that the complete modal spectrum is given by an integration over all the singularities of an appropriate one-dimensional Green's function, called the characteristic Green's function. In closed waveguides, only pole singularities are present and the resulting modal spectrum is purely discrete. In open regions, branch point singularities also occur, and the integration for the modal spectrum must be performed on the proper branch, the one corresponding to decaying waves at infinity. The residues of the poles, if any, on the proper branch then correspond to true discrete modes, while the branch cut integration(s) yields the continuous spectrum.

Also, pole singularities are often present on the wrong branch, and in the past their presence, if noted, was ignored since these poles did not contribute to the spectrum and were unphysical in that they corresponded to growing waves at infinity. However, Marcuvitz[14] recognized that these poles are in fact leaky waves, that under certain conditions some of them may contribute to the radiated field, and that they are valid only in certain suitably restricted regions.

The characteristic Green's function can be identified as the voltage (or current, depending on the problem) on a transmission line along one of the coordinates transverse to the axial direction of the waveguide (for example, the x direction in Fig. 1). Translated into microwave network terms, the singularities of this Green's function are actually the resonances of the appropriate transverse network representation of the structure. It is seen, therefore, that the transverse resonance procedure yields not only the true discrete modes of an open structure, as is well known, but is also the condition for leaky waves.

III. Perturbation Solution of the Resonance Equation

The general form of terminated transmission line which is representative of a transversely viewed waveguide is shown in Fig. 2.

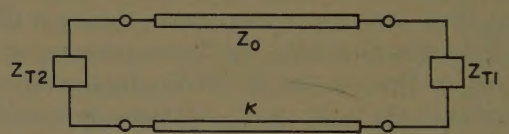


Fig. 2—Terminated transmission line representing a waveguide cross section.

The transverse wavenumbers κ corresponding to source-free fields in the waveguide are obtained from the solution of the resonance equation

$$\overleftrightarrow{Z}(\kappa) = \overleftarrow{Z}(\kappa) + \overrightarrow{Z}(\kappa) = 0$$

or

$$\overleftrightarrow{Y}(\kappa) = \overleftarrow{Y}(\kappa) + \overrightarrow{Y}(\kappa) = 0, \tag{6}$$

where $\overleftarrow{Z}$, $\overleftarrow{Y}$, $\overrightarrow{Z}$, $\overrightarrow{Y}$ are the impedances or admittances looking to the left and right, respectively, from some arbitrary reference plane. This resonance equation is usually a complex transcendental equation for which an exact solution cannot readily be obtained. Numerical methods or computing machines may be employed in specific cases but such numerical solutions are not of particular value in gaining insight into the general behavior of leaky structures.

Since for the class of leaky-wave antennas considered here the leaky wave can be regarded as a perturbation of a mode in a "closed" waveguide,[15] the transverse

[13] N. Marcuvitz, "Field representations in spherically stratified regions," *Commun. Pure and Appl. Math.*, vol. 4, pp. 284–293; August, 1951.

[14] N. Marcuvitz, "On field representations in terms of leaky modes or eigenmodes," IRE Trans. on Antennas and Propagation, vol. AP-4, pp. 192–194; July, 1956.

[15] A waveguide is regarded as "closed" when it does not radiate and consequently possesses real propagation wavenumbers. Reactive wall waveguides are included in this category.

wavenumber of the leaky wave can be written as

$$\kappa = \kappa_0 + \Delta\kappa, \tag{7}$$

where κ_0 is the wavenumber of the unperturbed mode and $\Delta\kappa$ is the perturbation to be determined. For homogeneous cross sections the resonance equation (6) may be written in terms of normalized impedances (admittances) by dividing through by the characteristic impedance (admittance). Expansion of the normalized form of (6) about the point $\kappa=\kappa_0$ to first order in $\Delta\kappa$ yields

$$\overleftrightarrow{Z'}(\kappa_0) + \left(\frac{d\overleftrightarrow{Z'}(\kappa)}{d\kappa}\right)_{\kappa=\kappa_0} \Delta\kappa \approx 0,$$

so that

$$\Delta\kappa \approx -\frac{\overleftrightarrow{Z'}(\kappa_0)}{\left(\frac{d\overleftrightarrow{Z'}(\kappa)}{d\kappa}\right)_{\kappa=\kappa_0}}, \tag{8a}$$

or

$$\Delta\kappa = -\frac{\overleftrightarrow{Y'}(\kappa_0)}{\left(\frac{d\overleftrightarrow{Y'}(\kappa)}{d\kappa}\right)_{\kappa=\kappa_0}}, \tag{8b}$$

where

$$\overleftrightarrow{Z'} = \frac{\overleftrightarrow{Z}}{Z_0} \quad \text{and} \quad \overleftrightarrow{Y'} = \frac{\overleftrightarrow{Y}}{Y_0}.$$

The reference plane for the resonance equation is normally chosen at the slot of the leaky guide as a matter of convenience. For larger perturbations, iteration procedures may be employed but the results of such iterations must be checked for convergence in each case.

For a leaky rectangular waveguide such as that shown in Fig. 3, the transverse transmission line representation is along the x direction and is of the type shown in Fig. 2, but with $Z_{T2}=0$. The length of the transmission line is, of course, equal to the guide width a. The propagation wavenumber along the z direction for this type of leaky guide is given by

$$k_z = \sqrt{k^2 - k_y^2 - \kappa^2}, \tag{9}$$

where $k=2\pi/\lambda$ and k_y is the propagation wavenumber in the y direction. When (7) is substituted into (9) and the term involving $(\Delta\kappa)^2$ is neglected,

$$k_z \approx \sqrt{k^2 - k_y^2 - \kappa_0^2 - 2\kappa_0\Delta\kappa} \tag{10a}$$

$$\approx \frac{2\pi}{\lambda_{g0}}\left(1 - \frac{\kappa_0\lambda_{g0}^2}{4\pi^2}\Delta\kappa\right), \tag{10b}$$

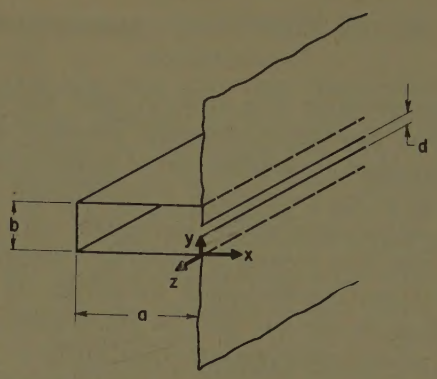


Fig. 3—Slitted rectangular waveguide.

where

$$\lambda_{g0} = \frac{2\pi}{\sqrt{k^2 - k_y^2 - \kappa_0^2}}$$

is the guide wavelength along z of the unperturbed mode. Then the real and imaginary parts of the propagation constant along z are

$$\frac{\beta}{k} = \frac{\lambda}{\lambda_g} \approx \frac{\lambda}{\lambda_{g0}}\left(1 - \frac{\kappa_0\lambda_{g0}^2}{4\pi^2}\Delta\kappa_r\right)$$

$$\alpha\lambda \approx \frac{\lambda\lambda_{g0}}{2\pi}\kappa_0\Delta\kappa_i, \tag{11}$$

where we have set $k_z=\beta-j\alpha$ and $\Delta\kappa=\Delta\kappa_r+j\Delta\kappa_i$. Approximations (10) and (11) should be used only when the frequency is well above the cut-off frequency of the unperturbed mode. When the cut-off frequency is approached, (9) must be used without approximation.

The perturbation $\Delta\kappa$ of the transverse wavenumber for a rectangular leaky waveguide is obtained from (8) and is

$$\Delta\kappa \approx -\frac{Z_T'(\kappa_0) + j\tan\kappa_0 a}{\left(\frac{dZ_T'}{d\kappa}\right)_{\kappa=\kappa_0} + ja\sec^2\kappa_0 a}, \tag{12a}$$

or

$$\Delta\kappa \approx -\frac{Y_T'(\kappa_0) - j\cot\kappa_0 a}{\left(\frac{dY_T'}{d\kappa}\right)_{\kappa=\kappa_0} + ja\csc^2\kappa_0 a}, \tag{12b}$$

where the terminating admittance $Y_T'=1/Z_T'$. The derivative of the terminating impedance or admittance, appearing in the denominators of (12), can frequently be neglected for nonresonant terminations so that (12) reduces to the simplified expressions

$$\Delta\kappa \approx \frac{j}{a}\frac{Z_T'(\kappa_0) + j\tan\kappa_0 a}{\sec^2\kappa_0 a}, \tag{13a}$$

or

$$\Delta\kappa \approx \frac{j}{a}\frac{Y_T'(\kappa_0) - j\cot\kappa_0 a}{\csc^2\kappa_0 a}. \tag{13b}$$

IV. Modes for Transverse Transmission Line Descriptions

The representation of a uniform open waveguide by a transverse equivalent network involves first, the introduction of an appropriate set of modes, and second, the solution of an appropriate discontinuity problem. When an open rectangular waveguide such as that shown in Fig. 3 is viewed transversely along x it can be regarded as a parallel-plate waveguide of height b terminated in an electric wall, or short circuit, at the plane $x=-a$ and radiating into a half space at the plane $x=0$. The modal representation appropriate to the description of such a structure in the region $-a<x<0$ by a transmission line along x is therefore that in terms of characteristic modes of a parallel-plate waveguide.

The desired modal representation of the transverse (to x) fields is[16]

$$\boldsymbol{E}_t(x, y, z) = \sum_i V_i(x)\boldsymbol{e}_i(y, z) \tag{14a}$$

$$\boldsymbol{H}_t(x, y, z) = \sum_i I_i(x)\boldsymbol{h}_i(y, z). \tag{14b}$$

These electric and magnetic fields satisfy the transverse part of the Maxwell field equations. When (14) is substituted into the transverse field equations, a separation may be effected in which the x dependent quantities V and I satisfy transmission line equations, and the mode functions $\boldsymbol{e}$ and $\boldsymbol{h}$ satisfy coupled second-order equations which constitute (together with appropriate boundary conditions) the eigenvalue problem for the modes. When $\boldsymbol{e}$ and $\boldsymbol{h}$ satisfy appropriate orthogonality conditions, these coupled equations can be transformed into an uncoupled pair of second-order equations.

It is known that in a parallel-plate waveguide the modes separate into two types, and these can be chosen in a variety of ways. In particular, the mode functions may be chosen to satisfy the relationship $\boldsymbol{e}_i=\boldsymbol{h}_i\times\mathbf{x}_0$. If this is done the resulting modes are the usual E and H modes along x characterized by $H_x=0$ and $E_x=0$, respectively.[16] This type of representation suffers from a disadvantage in connection with leaky-wave structures since the slot in the guide wall produces coupling between the two types of mode, and thereby complicates the treatment of the discontinuity problem. However, if an alternative representation is used such that for one set $E_z=0$ and for the other $H_z=0$, no coupling between these two mode types results *since the slot is uniform in the z direction*. These alternative modes are designated as H-type and E-type modes, respectively. The use of this alternative representation effects the decomposition of a vector electromagnetic diffraction problem into two scalar problems with a resultant simplification in the treatment of the discontinuity.

[16] The notation employed here is that given by N. Marcuvitz, "Waveguide Handbook," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 10, ch. 1; 1951.

It follows that for these E-type and H-type modes the mode functions are characterized as follows:

$$\text{H-type: } e_{zi} = 0$$

$$\text{E-type: } h_{zi} = 0.$$

The general form of the vector mode functions which result from this characterization, and their orthogonality properties, are presented elsewhere.[17] For parallel plate geometry, the explicit form of the E-type mode functions, which are designated by a prime, is[18]

$$\boldsymbol{e}_n' = \sqrt{\frac{2}{b}}\sin\frac{n\pi y}{b}\,\mathbf{z}_0 - j\sqrt{\frac{2}{b}}\,\frac{n\pi}{b}\,\frac{k_z}{k^2-k_z^2}\cos\frac{n\pi y}{b}\,\mathbf{y}_0 \tag{15a}$$

$$h_n' = \sqrt{\frac{2}{b}}\sin\frac{n\pi y}{b}\,y_0 \tag{15b}$$

where

$$\kappa_n^2 = k^2 - \left(\frac{n\pi}{b}\right)^2 - k_z^2$$

$$n = 1, 2, 3, \cdots, -\infty < k_z < \infty$$

and the characteristic impedance Z_n' is given by

$$Z_n' = \frac{1}{Y_n'} = \frac{k^2-k_z^2}{\omega\epsilon\kappa_n}. \tag{16}$$

The explicit form of the H-type mode functions, which are designated by a double prime, becomes

$$\boldsymbol{e}_n'' = -\sqrt{\frac{\epsilon_n}{b}}\cos\frac{n\pi y}{b}\,\mathbf{y}_0 \tag{17a}$$

$$\boldsymbol{h}_n'' = \sqrt{\frac{\epsilon_n}{b}}\cos\frac{n\pi y}{b}\,\mathbf{z}_0 + j\sqrt{\frac{2}{b}}\,\frac{n\pi}{b}\,\frac{k_z}{k^2-k_z^2}\sin\frac{n\pi y}{b}\,\mathbf{y}_0 \tag{17b}$$

where

$$\kappa_n^2 = k^2 - \left(\frac{n\pi}{b}\right)^2 - k_z^2$$

$$\epsilon_n = \begin{cases} 1, & n = 0 \\ 2, & n > 0 \end{cases}$$

$$n = 0, 1, 2, 3, \cdots, -\infty < k_z < \infty,$$

and the characteristic impedance Z_n'' is

$$Z_n'' = \frac{1}{Y_n''} = \frac{\omega\mu\kappa_n}{k^2-k_z^2}. \tag{18}$$

[17] L. O. Goldstone and A. A. Oliner, "Leaky Wave Antennas I: Rectangular Waveguides," Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., Rept. R-606-57, PIB-534; August 11, 1957.

[18] The exponential dependence on z has been suppressed since it is common to all the mode functions.

It is significant to note that for $k_z=0$ (no variation along z) the E-type mode functions and characteristic impedance become identical with those for the conventional H modes along x ($E_z=0$) and the H-type mode functions and characteristic impedance become identical with those for the familiar E modes along x ($H_z=0$). Thus, when there is no variation along z, the usual E and H modes of a parallel-plate waveguide are simultaneously H-type modes and E-type modes, respectively. This feature will be exploited in the next section to obtain the circuit parameters of the lumped terminations for the transverse transmission line representations of certain uniform leaky structures.

Although the modes given in (15) and (17) are the most convenient for the circuit representation of such discontinuity structures as uniform infinite slots, they are not appropriate for the treatment of periodic discontinuities which occur in the case of periodic leaky waveguides (see Fig. 4). In such cases, the modes must be characterized by a discrete infinite set of wavenumbers along z corresponding to the spatial harmonics of the periodic discontinuity. The set of mode functions appropriate to periodic discontinuities is given elsewhere.[17]

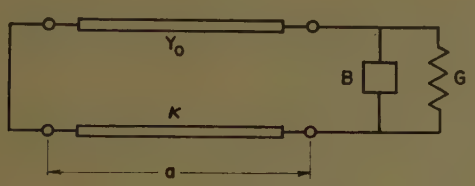


Fig. 4—Transverse equivalent network for uniform leaky rectangular waveguide.

V. Transverse Equivalent Networks

The modes which have been discussed in the preceding section provide a transverse transmission line representation of leaky rectangular waveguides. To complete the transverse network description it is necessary to obtain a lumped network characterization of the discontinuity represented by the slot or apertures in the waveguide wall and the external region into which the guide radiates.

In this section transverse network representations will be obtained for several uniform leaky waveguides of the type shown in Fig. 3, and for a leaky waveguide in which one of the narrow faces contains a periodic array of circular apertures.

A. Uniform Leaky Waveguides

There are two leaky waves which are of primary interest in connection with uniform slotted waveguides. The first such leaky wave is one for which the field distribution corresponds to that of the H_{10} mode in a closed waveguide; the second is a wave which corresponds to the E_{11} mode (see diagram 4 of the Appendix). Each of these poses a different discontinuity problem so that different transverse networks must be obtained for each case.

In the first case, the transverse discontinuity problem involves an incident H-type parallel-plate mode with no variation in the y direction ($n=0$) but with arbitrary periodicity in the z direction. Because of the uniformity of the structure along z this discontinuity problem involves only H-type modes. Since the conditions here are those corresponding to the H_{10} mode in a closed waveguide, the frequency is such that $\lambda>2b$ and, therefore, all higher modes ($n>0$) are nonpropagating. In the second case, the incident wave is an E-type mode with a y dependence given by $\sin \pi y/b$ and with arbitrary periodicity in the z direction. Only E-type modes are involved, and since this case corresponds to the E_{11} mode in a closed waveguide the frequency is such that $2b>\lambda>b$. Consequently, all the higher modes ($n>1$) are nonpropagating. The required mode functions are those given in (15) or (17).

Since the basic discontinuity problem in both cases is that of a parallel-plate waveguide radiating into a half space, a general stationary expression for the normalized admittance of the lumped network which is valid for both cases can be written as follows:

$$\frac{Y}{Y_0}=\sum_n{}'\frac{Y_n|V_n|^2}{Y_0|V_0|^2}+\int_{ap}dy\int_{ap}dy'\,\frac{\mathbf{x}_0\times E_t(y')\cdot\mathcal{Y}_h\cdot\mathbf{x}_0\times E_t(y)}{Y_0|V_0|^2},\quad(19)$$

where Y_0 and V_0 are the characteristic admittance and voltage of the propagating mode, the Y_n and V_n are the characteristic admittances and voltages of the nonpropagating modes, and $\mathcal{Y}_h$ is the two-dimensional half-space dyadic Green's function. The prime on the summation sign in (19) indicates that the sum is taken over the nonpropagating modes only.

The first term on the right-hand side of (19) is purely imaginary and represents the contribution to the susceptance due to the stored energy in the internal region $x<0$. In writing (19) the assumption has been made that none of the nonpropagating modes "see" the back wall of the guide at $x=-a$. Since the nonpropagating modes decay rapidly away from the slot, this assumption is normally valid. However, if the back wall is quite close, the assumption may be incorrect for a few of the lowest order nonpropagating modes. Under these circumstances, it is possible to effect a simple correction to take into account the interaction with the back wall. This correction consists of replacing the characteristic admittances in the first term on the right-hand side of (19) by the actual admittances seen by those modes which are affected by the back wall. These actual admittances are just the input admittances of short-circuited lengths of cut-off transmission lines and are

$$\overleftarrow{Y_n}=Y_n\coth|\kappa_n|a,$$

where Y_n is the characteristic admittance of the mode and κ_n is its wavenumber for propagation along x. For all of the cases considered in this report the correction is negligible for all the nonpropagating modes.

The second term of (19) is complex. It represents the conductance due to the radiated energy and the contribution to the susceptance due to the stored energy in the external region $x>0$. This term, of course, is in no way affected by the presence of the back wall.

For the two cases being considered here the conductances and susceptances are known for the case $k_z=0$. These known results can be readily extended for arbitrary k_z. The required modification can be obtained by noting that when $k_z=0$ the field components E_z and H_z satisfy

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+k^2\right)\frac{E_z}{H_z}=0,$$

and when $k_z\neq 0$ they satisfy

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+k^2-k_z^2\right)\frac{E_z}{H_z}=0.$$

Since the boundary conditions on E_z and H_z are independent of k_z, the solutions for $k_z\neq 0$ can be obtained directly from those for $k_z=0$ by replacing k in the latter by $\sqrt{k^2-k_z^2}$. The z components of the fields of the propagating modes in each case can be represented as

$$H_z=[I''_{\text{inc}}(x)+I''_{\text{refl}}(x)]h''_z(y) \quad \text{for the H-type mode,}$$

$$E_z=[V'_{\text{inc}}(x)+V'_{\text{refl}}(x)]e'_z(y) \quad \text{for the E-type mode.}$$

If (voltage) reflection coefficients are defined so that

$$\frac{H_{z\,\text{refl}}}{H_{z\,\text{inc}}}=-\Gamma''=\frac{I''_{\text{refl}}}{I''_{\text{inc}}}$$

and

$$\frac{E_{z\,\text{refl}}}{E_{z\,\text{inc}}}=\Gamma'=\frac{V'_{\text{refl}}}{V'_{\text{inc}}},$$

it can be seen that the reflection coefficients for the case $k_z\neq 0$ must be related to those for $k_z=0$ in the same manner as the field components. Since normalized admittances are related to reflection coefficients by

$$\frac{Y}{Y_0}=\frac{1-\Gamma}{1+\Gamma},$$

the replacement of k by $\sqrt{k^2-k_z^2}$ is also the required modification for the normalized admittances.

Since the field in the vicinity of the slot for $x<0$ for both the H-type and E-type modes depends primarily on the local geometry and not on the nature of the region $x>0$, the "internal" susceptances, which are represented by the first term of (19), can be approximated by one-half the capacitive and inductive susceptances, respectively, of a slit aperture in an infinite parallel-plate waveguide. When the foregoing replacement of k by $\sqrt{k^2-k_z^2}$ is made, the internal susceptances are given approximately, for the H-type mode, by[19]

$$B'_{\text{int}}=\frac{B_{\text{int}}}{Y_0}\approx\frac{\kappa b}{\pi}\ln\left(\csc\frac{\pi d}{2b}\right), \tag{20}$$

with

$$\kappa=\sqrt{k^2-k_z^2},$$

and for the E-type mode by[20]

$$B'_{\text{int}}=\frac{B_{\text{int}}}{Y_0}\approx-\frac{\pi}{\kappa b}\cot^2\frac{\pi d}{2b}, \tag{21}$$

$$\kappa=\sqrt{k^2-(\pi/b)^2-k_z^2}.$$

The external susceptances and conductances for both the H-type and E-type modes are obtained by suitable modifications of results given in the "Waveguide Handbook" (WGH).[21,22] These WGH expressions correspond to the special case for which the waveguide height is equal to the height of the slit. When the slit height is less than the waveguide height these expressions are modified in the following manner. One first recognizes that the power per unit length radiated into a half space from a slit is independent of the height of the guide feeding the slit, provided that the slit height and the slit electric field remain unchanged. If the subscripts 1 and 2 denote the cases for which the slit and guide heights are equal and unequal, respectively, we may then write

$$Y_1V_1^2=Y_2V_2^2,$$

where Y is the input admittance of the slit and V is the slit voltage, given by

$$V=\int_{\text{slit}} n\times E\cdot h\,dy,$$

where E is the slit aperture electric field, n is the normal to the slit, and h is the dominant mode function of the feeding waveguide. The power equality relation can be solved for Y_2 and appropriately normalized to yield

$$\frac{Y_2}{Y_{02}}=\frac{Y_1}{Y_{01}}\left(\frac{Y_{01}}{Y_{02}}\right)\left(\frac{V_1}{V_2}\right)^2.$$

Quantity Y_1/Y_{01} is the normalized input admittance given in the "Waveguide Handbook."[21,22] The ratio of characteristic admittances is

$$\frac{Y_{01}}{Y_{02}}=\begin{cases}\dfrac{\kappa_2}{\kappa_1} & \text{for H-type modes}\\[2ex] \dfrac{\kappa_1}{\kappa_2} & \text{for E-type modes,}\end{cases}$$

[19] Marcuvitz, *op. cit.*, footnote 16, p. 218, (2a).
[20] *Ibid.* Only the $k_z=0$ case is available, and is given by (1b) on p. 221.
[21] *Ibid.*, p. 184, (1b) and (2b).
[22] *Ibid.* The $k_z=0$ case is given by (2b) and (3b) on p. 191.

while the slit electric field, occurring in the expressions for the voltages, is either

$$E(y) = \begin{cases} 1 & \text{for H-type modes} \\ \cos\dfrac{\pi y}{d} & \text{for E-type modes,} \end{cases}$$

where y is measured from the center of the slit. The amplitude of the field is of no consequence since the voltage terms occur in a ratio.

Employing the procedure outlined above, the external conductances and susceptances become, for the H-type mode,

$$G' = \frac{G}{Y_0} \approx \frac{\kappa b}{2} \tag{22a}$$

$$B'_{\text{ext}} = \frac{B_{\text{ext}}}{Y_0} \approx \frac{\kappa b}{\pi} \ln \frac{\pi e}{\gamma \kappa d}, \tag{22b}$$

where $e=2.718$, $\gamma=1.781$, $\kappa=\sqrt{k^2-k_z^2}$, and for the E-type mode,

$$G' = \frac{G}{Y_0} \approx 0.285 \left[\frac{\pi^2 b(1-(d/b)^2)^2}{16d\cos^2(\pi d/2b)}\right]\frac{2\pi}{\kappa d} \tag{23a}$$

$$B' = \frac{B}{Y_0} \approx -0.156 \left[\frac{\pi^2 b(1-(d/b)^2)^2}{16d\cos^2(\pi d/2b)}\right]\frac{2\pi}{\kappa d}, \tag{23b}$$

with

$$\kappa = \sqrt{k^2 - (\pi/b)^2 - k_z^2}.$$

The complete transverse equivalent network for a uniform leaky rectangular waveguide has the form shown in Fig. 4 for both modes, where B represents $(B_{\text{int}}+B_{\text{ext}})$. For the H-type mode B is capacitive, while for the E-type mode it is inductive. It might also be noted that the equivalent network of Fig. 4 with

$$\frac{G}{Y_0} = 1$$

and

$$\frac{B}{Y_0} = \frac{2\kappa b}{\pi} \ln\left(\csc\frac{\pi d}{2b}\right)$$

represents the long-slot array treated by Honey[10] (diagram 2 of the Appendix).

In the case of the transverse network for the E-type leaky wave, the use of E-type modes is particularly significant. If conventional E and H modes were used, the transverse network would consist of two transmission lines, one for an E mode and one for an H mode, which are coupled by the slot. Furthermore, the parameters of the coupling network are not known and would have to be evaluated. With the use of E-type modes, only a single transmission line is required, and the discontinuity admittance is obtainable immediately from a known result upon replacement of k by $\sqrt{k^2-k_z^2}$, an operation whose justification also requires the use of E-type or H-type modes.

B. Periodic Leaky Waveguides

An example of the class of periodic leaky waveguides is the one shown in Fig. 5. The apertures are circular holes of diameter d. The principal leaky wave in this structure can be regarded as a perturbation on the dominant H_{10} mode in the closed guide. Although structures such as the one shown in Fig. 5 differ considerably, in some respects, from the uniform leaky guides, when the aperture spacing is small enough ($c<\lambda/2$) their leaky-wave behavior is similar to that of the uniform structure.

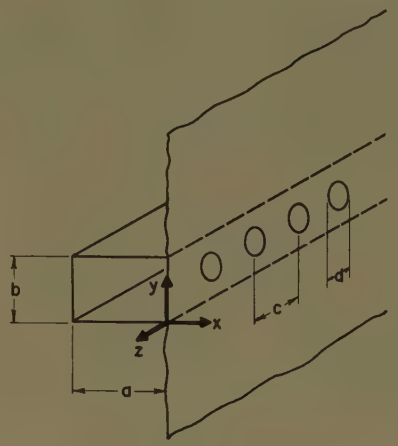


Fig. 5—Periodic leaky waveguide.

While the transverse discontinuity admittance for this periodic structure is not available, it can be derived relatively easily by appealing to a "small aperture" procedure. The results are therefore subject to the restrictions $kd<1$ and $d\ll b$, but are quite accurate otherwise. Under these circumstances the electric field established in any aperture by an incident field is the same as that which would be produced in this aperture in the absence of all the others.

The incident wave for this discontinuity problem is taken to be an H-type parallel-plate mode with no variation in the y direction and with a z dependence exp $(-jk_{z0}z)$, where $k_{z0}(=\sqrt{k^2-(\pi/a)^2})$ is the wavenumber of the unperturbed H_{10} mode, *i.e.*, in the closed guide. The complete set of H-type modes for a periodically-loaded parallel-plate waveguide have been given elsewhere.[17] The incident mode here can be identified as the lowest member $(m=n=0)$ of that set. Because of the small aperture approximation only H-type modes are excited by the array, and since the elements of the array are assumed to be spaced less than a half free-space wavelength apart all modes except the lowest are nonpropagating in the x direction.

From the small aperture point of view, the aperture fields and the stored energy in the vicinity of the apertures are relatively independent of the surrounding waveguide geometry. For this reason, the susceptance portion of the admittance may be well approximated

by the susceptance corresponding to the situation where the region $x>0$ is a parallel-plate waveguide rather than a half space. To first order, this susceptance can be expressed in unnormalized fashion as[23]

$$B \approx -\frac{1}{\omega\mu M |h|^2}, \tag{24}$$

where $M=d^3/6$ is the magnetic polarizability of a circular aperture and $|h|=1/\sqrt{bc}$ is the absolute value of the mode function for the incident mode $(m=n=0)$. The normalized susceptance can therefore be written as

$$B' = \frac{B}{Y_0} \approx \frac{6bc}{\kappa d^3}, \tag{25}$$

where

$$Y_0 = \kappa/\omega\mu, \qquad \kappa = \kappa_0 = \sqrt{k^2 - k_{z0}^2}.$$

The conductance may be calculated from

$$G = \frac{P}{|V^2|}, \tag{26}$$

where P is the real transverse power flow per unit cell and V is the modal voltage of the propagating mode at the aperture plane. Since the transverse power flow per unit cell within the guide must be equal to the outward radial power flow per unit cell in the half space, the numerator of (26) can be calculated from the far fields in the half space. The details of the calculation of the conductance are given elsewhere,[17] and it is shown that the expression for the unnormalized conductance is

$$G \approx \frac{b\kappa^2}{2\omega\mu}, \tag{27}$$

while that for the normalized conductance is

$$G' = \frac{G}{Y_0} \approx \frac{\kappa b}{2}. \tag{28}$$

The complete transverse equivalent network of the periodic leaky guide is shown in Fig. 6.

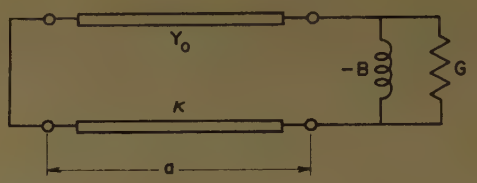


Fig. 6—Transverse equivalent network for the periodic leaky waveguide.

VI. Theoretical and Experimental Results

The leaky-wave propagation constants of the specific structures for which transverse equivalent networks have been derived in the preceding section are now readily obtained from the perturbation relations (11) and (13).

A. Narrow Slots and Small Apertures

For the case of the leaky waveguides shown in Figs. 3 and 5, where the slots and apertures are small $(d \ll b)$, the appropriate unperturbed structure is simply a completely closed rectangular waveguide of dimensions a by b. The unperturbed transverse wavenumber for the H-type waves in both the uniform and periodic waveguides and the E-type wave in the uniform guide is then

$$\kappa_0 = \pi/a. \tag{29}$$

The attenuation constants and guide wavelengths for these cases are given by

$$\alpha\lambda \approx \frac{\lambda\lambda_{g0}}{2a^2} R'\left(\frac{\pi}{a}\right) \tag{30a}$$

$$\frac{\lambda}{\lambda_g} \approx \frac{\lambda}{\lambda_{g0}}\left(1 \pm \frac{\lambda_{g0}^2}{4\pi a^2} X'\left(\frac{\pi}{a}\right)\right). \tag{30b}$$

For the H-type wave in both the uniform and periodic structures

$$\lambda_{g0} = \frac{\lambda}{\sqrt{1-\left(\frac{\lambda}{2a}\right)^2}},$$

and the positive and negative signs in (30b) apply to the periodic and uniform guides, respectively. For the E-type wave in the uniform structure

$$\lambda_{g0} = \frac{\lambda}{\sqrt{1-\left(\frac{\lambda}{2a}\right)^2-\left(\frac{\lambda}{2b}\right)^2}},$$

and the positive sign in (30b) is to be used. The quantities R' and X' are given by

$$R' = \frac{\frac{G}{Y_0}}{\left(\frac{G}{Y_0}\right)^2+\left(\frac{B}{Y_0}\right)^2}$$

$$X' = \frac{\frac{B}{Y_0}}{\left(\frac{G}{Y_0}\right)^2+\left(\frac{B}{Y_0}\right)^2},$$

where B/Y_0 and G/Y_0 are the appropriate parameters for each case as given in the preceding section. The approximate expressions for attenuation constants and guide wavelengths given by (30) for these cases hold only when the frequency is well above the cut-off frequency of the corresponding unperturbed mode. In the

[23] N. Marcuvitz, "Waveguide Circuit Theory: Coupling of Waveguides by Small Apertures," Microwave Res. Inst., Polytechnic Inst. of Brooklyn, N. Y., Rept. R-157-47, PIB-106; 1947.

case of the H-type waves in the uniform and periodic guides these expressions are therefore valid only if $\lambda \ll 2a$. In the case of the E-type wave in uniform guide they are valid only if $\lambda \ll 2ab/\sqrt{a^2+b^2}$. When the above conditions are not fulfilled the desired propagation constants must be computed from (9).

The approximate expressions for guide wavelength and attenuation constant given in (30) also apply to the long-slot array treated by Honey (see diagram 2 of the Appendix) for the leaky wave which he designates as the "π mode."[10] To check the usefulness of the perturbation procedure, calculations were made for the Stanford Research Institute long-slot array using (30). A comparison of these values with the results obtained by Honey from an exact numerical solution of the resonance equation is shown in Fig. 7. The perturbation calculations yield almost exact agreement for λ/λ_g values, and although the agreement for the αa values is not as good as that for λ/λ_g it can be seen to be quite satisfactory.

Calculations were also made for the periodic leaky waveguide. A comparison of measured and calculated values of guide wavelength for this structure is shown in Fig. 8. The attenuation constant of the experimental structure was too small to be measured.

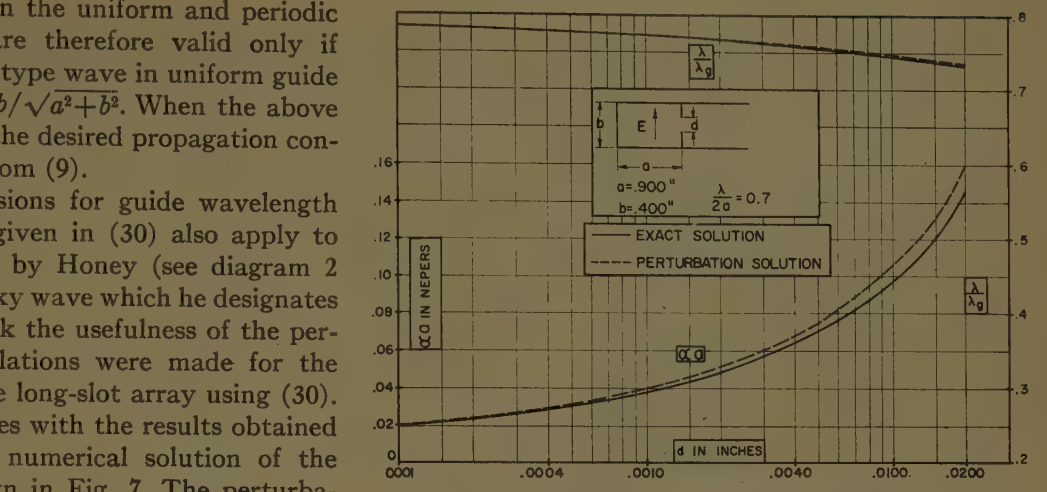


Fig. 7—Propagation characteristics vs slot width for leaky wave on long slot array.

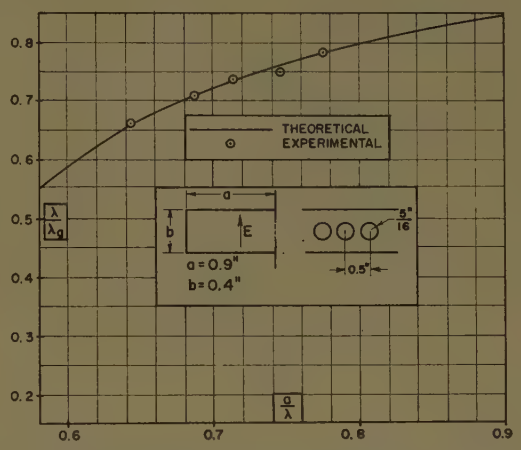


Fig. 8—Guide wavelength of H-type leaky wave in rectangular waveguide with circular hole array.

B. Wide Slots

When the slot in the uniform leaky guide of Fig. 3 is not narrow the structure cannot simply be regarded as a perturbation on an ordinary closed waveguide. Nevertheless, very good results can be obtained by a modified perturbation procedure.

The most extreme case is that of the so-called channel waveguide in which the slot width is equal to the narrow dimension of the guide (see diagram 3 of the Appendix). A possible unperturbed structure in this case is a rectangular waveguide with a magnetic wall for one of the narrow faces. The corresponding unperturbed wavenumber is then $\kappa_0 = \pi/2a$. For a very flat channel guide, for which $a \gg b$, this simple perturbation procedure yields satisfactory results since for this situation the aperture admittance is small and can therefore be treated as a small perturbation on an open circuit.

For the usual waveguide dimensions, however, this is not the case. Since the phase part of the desired propagation constant depends primarily on the energy storage property (susceptance) of the discontinuity, the first step in the perturbation procedure consists of formulating a resonance problem, involving only the susceptance, in the following form:

$$\cot \kappa_0 a = B'(\kappa_0). \tag{31}$$

This equation (for real κ_0) may be readily solved to any desired degree of accuracy by a successive approximation or graphical procedure, or may be solved approximately by perturbing about an open-circuit termination (magnetic wall). The perturbation approach yields the solution

$$\kappa_0 \approx \frac{\pi}{2a} - \frac{1}{a} B'\left(\frac{\pi}{2a}\right). \tag{32}$$

The solution of (31) (either numerical or perturbation) is then used in (13b) and yields

$$\Delta\kappa = \frac{j}{a} \frac{G'(\kappa_0)}{\csc^2 \kappa_0 a}. \tag{33}$$

The attenuation constant and guide wavelength of the H-type leaky wave in the channel guide are then obtained as

$$\frac{\beta}{k} = \frac{\lambda}{\lambda_g} \approx \sqrt{1 - \frac{\lambda^2 \kappa_0^2}{4\pi^2}}, \tag{34a}$$

$$\alpha\lambda \approx \left(\frac{\lambda}{a}\right) \frac{\lambda_g \kappa_0}{2\pi} \frac{G'(\kappa_0)}{\csc^2 \kappa_0 a}. \tag{34b}$$

In Figs. 9 and 10 theoretical curves computed from (34) for λ/λ_g and $\alpha\lambda$ for the "lowest" H-type leaky wave in a channel guide are compared with a set of measured results.

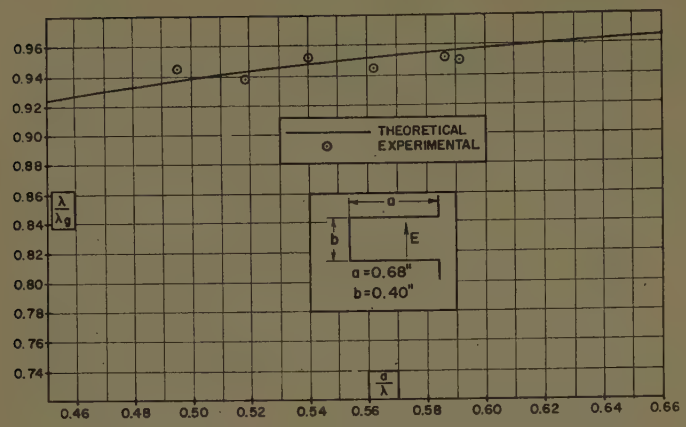


Fig. 9—Guide wavelength of H-type leaky wave in channel waveguide.

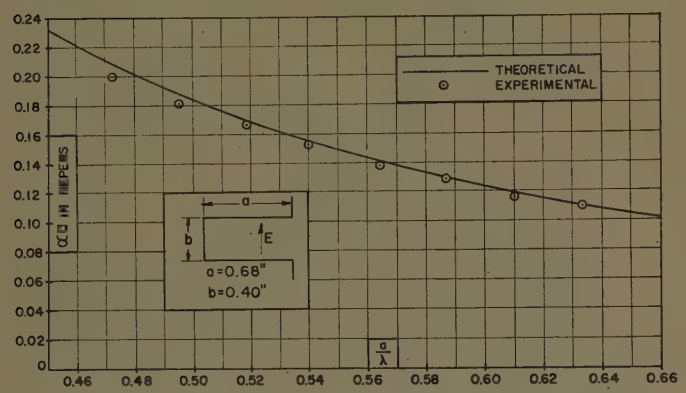


Fig. 10—Attenuation constant of H-type leaky wave in channel waveguide.

In the case of the E-type wave in the uniform guide when the slot is wide but does not occupy the entire narrow face of the guide, a procedure similar to that above yields very good results. The first step again consists of formulating a resonance problem using only the susceptance, as follows,

$$\tan \kappa_0 a = -\frac{1}{B'(\kappa_0)}. \tag{35}$$

The above equation for real κ_0 can be solved numerically or by perturbing about a short-circuit termination rather than an open-circuit termination as in the case of the channel guide. The perturbation procedure yields the solution

$$\kappa_0 = \frac{\pi}{a} - \frac{1}{a}\frac{1}{B'(\pi/a)}. \tag{36}$$

The solution of (35) is then used in (13a) to obtain the result

$$\Delta\kappa = \frac{j}{a}\frac{R'(\kappa_0) + jX'(\kappa_0) + j\tan \kappa_0 a}{\sec^2 \kappa_0 a}. \tag{37}$$

The guide wavelength and attenuation constant in this case are then given by

$$\frac{\lambda}{\lambda_g} \approx \frac{\lambda}{\lambda_{g0}}\left\{1 + (\kappa_0 a)\left(\frac{\lambda_{g0}}{2\pi a}\right)^2 \frac{(X'(\kappa_0) + \tan \kappa_0 a)}{\sec^2 \kappa_0 a}\right\} \tag{38a}$$

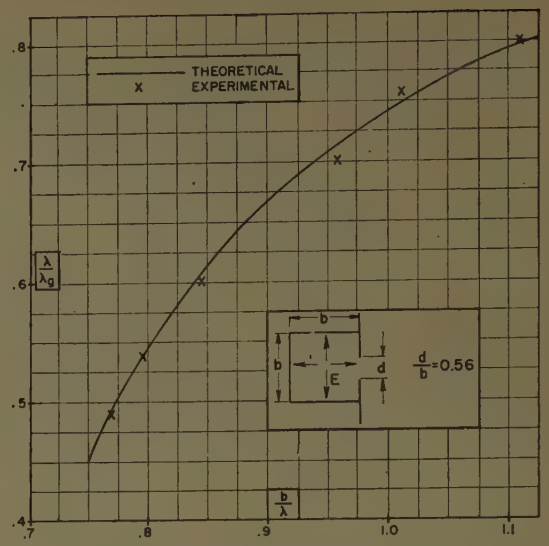


Fig. 11—Guide wavelength of E-type leaky wave in square waveguide.

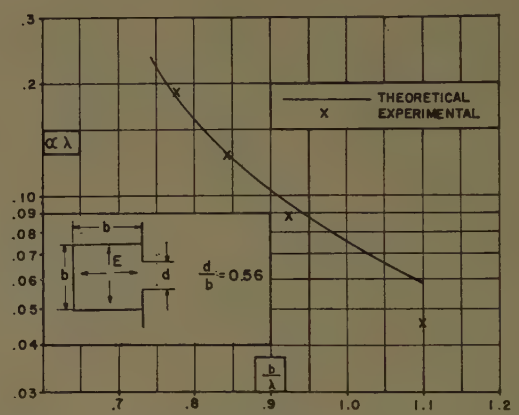


Fig. 12—Attenuation constant of E-type leaky wave in square waveguide.

$$\alpha\lambda \approx \frac{1}{2\pi}\left(\frac{\lambda_{g0}}{\lambda}\right)\left(\frac{\lambda}{a}\right)^2 \frac{(\kappa_0 a)R'(\kappa_0)}{\sec^2 \kappa_0 a}, \tag{38b}$$

where

$$\lambda_{g0} = \frac{\lambda}{\sqrt{1 - \left(\frac{\lambda}{2b}\right)^2 - \left(\frac{\lambda\kappa_0}{2\pi}\right)^2}}.$$

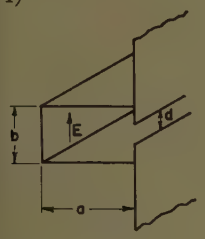
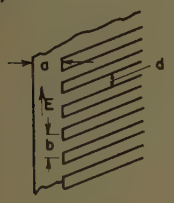
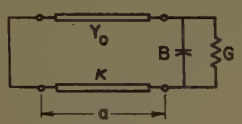
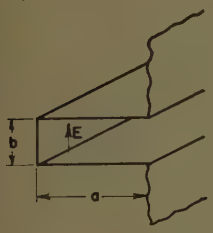
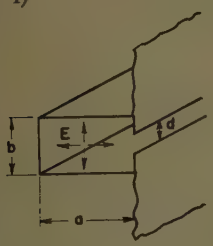
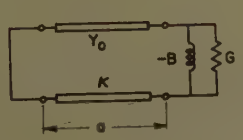
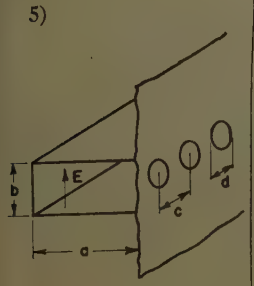
Theoretical curves computed from (38) for a square waveguide with $d/b = 0.56$ are compared in Figs. 11 and 12 with experimental results published by Hines, Rumsey, and Walter.[8]

In all of the above cases, the agreement between the measured data and the theoretical values can be observed to be very satisfactory.

Acknowledgment

The authors wish to thank Prof. N. Marcuvitz for helpful discussions in the early stages of this work. They also wish to express their appreciation to Mrs. A. Lotito for performing the numerical computations, and to S. Chertoff for taking the measurements.

Appendix
Tabulation of Theoretical Results

Structure	Equivalent Network	Attenuation Constant and Guide Wavelength	Circuit Parameters
1) 		$\alpha\lambda = \dfrac{\lambda\lambda_{g0}}{2a^2} R'$ $\dfrac{\lambda}{\lambda_g} = \dfrac{\lambda}{\lambda_{g0}}\left(1 - \dfrac{\lambda_{g0}^2}{4\pi a^2} X'\right)$ $R' = \dfrac{G'}{G'^2 + B'^2}, \; X' = \dfrac{B'}{G'^2 + B'^2}$ $\lambda_{g0} = \dfrac{\lambda}{\sqrt{1 - (\lambda/2a)^2}}$	$\dfrac{G}{Y_0} = G' = \dfrac{\pi b}{2a}$ $\dfrac{B}{Y_0} = B' = \dfrac{b}{a} \ln\left(\csc\dfrac{\pi d}{2b}\right) + \dfrac{b}{a} \ln\dfrac{ae}{\gamma d}$ $e = 2.718, \gamma = 1.781$
2) 		Same as above for structure 1)	$\dfrac{G}{Y_0} = G' = 1$ $\dfrac{B}{Y_0} = B' = \dfrac{b}{a} \ln\left(\csc\dfrac{\pi d}{2b}\right)$
3) 		$\alpha\lambda = \left(\dfrac{\lambda}{a}\right) \dfrac{\lambda_g \kappa_0}{2\pi} \dfrac{G'}{\csc^2 \kappa_0 a}$ $\dfrac{\lambda}{\lambda_g} = \sqrt{1 - \left(\dfrac{\lambda\kappa_0}{2\pi}\right)^2}$	$\dfrac{G}{Y_0} = G' = \dfrac{\kappa_0 b}{2}$ $\dfrac{B}{Y_0} = B' = \dfrac{\kappa_0 b}{\pi} \ln\dfrac{\pi e}{\gamma \kappa_0 b}$ $\kappa_0 = \dfrac{\pi}{2a} - \dfrac{b}{2a^2} \ln\dfrac{2ae}{\gamma b}$
4) 		Case for which $d \ll b$: $\alpha\lambda = \dfrac{\lambda\lambda_{g0}}{2a^2} R'$ $\dfrac{\lambda}{\lambda_g} = \dfrac{\lambda}{\lambda_{g0}}\left(1 - \dfrac{\lambda_{g0}^2}{4\pi a^2} X'\right)$ $R' = \dfrac{G'}{G'^2 + B'^2}, \; X' = \dfrac{B'}{G'^2 + B'^2}$ $\lambda_{g0} = \dfrac{\lambda}{\sqrt{1 - (\lambda/2a)^2 - (\lambda/2b)^2}}$	$\dfrac{G}{Y_0} = G' = 0.570 \left[\dfrac{\pi^2 b (1 - (d/b)^2)^2}{16 d \cos^2(\pi d/2b)}\right] (a/d)$ $\dfrac{B}{Y_0} = B' = \dfrac{a}{b} \cot^2\dfrac{\pi d}{2b} + .312 \left[\dfrac{\pi^2 b (1 - (d/b)^2)^2}{16 d \cos^2(\pi d/2b)}\right] (a/d)$
		$\alpha\lambda = \dfrac{1}{2\pi}\left(\dfrac{\lambda_{g0}}{\lambda}\right)\left(\dfrac{\lambda}{a}\right)^2 \dfrac{\kappa_0 a R'}{\sec^2 \kappa_0 a}$ $\dfrac{\lambda}{\lambda_g} = \dfrac{\lambda}{\lambda_{g0}}\left\{1 + \kappa_0 a \left(\dfrac{\lambda_{g0}}{2\pi a}\right)^2 \left(\dfrac{X' + \tan\kappa_0 a}{\sec^2 \kappa_0 a}\right)\right\}$ $\lambda_{g0} = \dfrac{\lambda}{\sqrt{1 - (\lambda/2b)^2 - (\lambda\kappa_0/2\pi)^2}}$	$\dfrac{G}{Y_0} = G' = 0.570 \left[\dfrac{\pi^2 b (1 - (d/b)^2)^2}{16 d \cos^2(\pi d/2b)}\right] (\pi/\kappa_0 d)$ $\dfrac{B}{Y_0} = B' = \dfrac{\pi}{\kappa_0 b} \cot^2\dfrac{\pi d}{2b} + 0.312 \left[\dfrac{\pi^2 b (1 - (d/b)^2)^2}{16 d \cos^2(\pi d/2b)}\right] (\pi/\kappa_0 d)$ $\kappa_0 = \dfrac{\pi}{a} - \dfrac{1}{a} \dfrac{1}{B'(\pi/a)}$
5) 		$\alpha\lambda = \dfrac{\lambda\lambda_{g0} R'}{2a^2}$ $\dfrac{\lambda}{\lambda_g} = \dfrac{\lambda}{\lambda_{g0}}\left(1 + \dfrac{\lambda_{g0}^2 X'}{4\pi a^2}\right)$ $\lambda_{g0} = \dfrac{\lambda}{\sqrt{1 - (\lambda/2a)^2}}$	$\dfrac{G}{Y_0} = G' = \dfrac{\pi b}{2a}$ $\dfrac{B}{Y_0} = B' = \dfrac{6abc}{\pi d^2}$

A Flush-Mounted Leaky-Wave Antenna with Predictable Patterns*

R. C. HONEY†

Summary—**This paper describes the design and the measured performance of a large, flat antenna consisting of an inductive grid spaced over a conducting surface. The analysis employs the transverse-resonance method to determine the radiating properties of the structure. This analytical technique is shown to predict very accurately the amplitude and phase of the illumination along the aperture of the antenna.**

An antenna was built with an 18- by 24-inch aperture and tested over the frequency band from 7-to-13 kmc. The results of these tests confirm the theoretical predictions in every detail. A pencil beam from the antenna scans in the H-plane (perpendicular to the antenna) from 20° to 60° from the normal to the aperture as the frequency changes from 7-to-13 kmc. The H-plane beamwidth remains virtually constant over most of this band. The first H-plane sidelobe or shoulder is at least 29 db below the main lobe from 7-to-10 kmc, and at least 23 db below from 10-to-13 kmc. All H-plane sidelobes beyond three or four beamwidths on either side of the main lobe are at least 40 db below the main lobe everywhere in the 7-to-13 kmc band. At the design frequency the measured pattern agrees with the theoretical pattern within a fraction of a db down to 40 db below the peak of the main lobe, even though the gain of the antenna at this frequency is only 33 db.

I. Introduction

"LEAKY-WAVE" antennas may be briefly described as transmission lines that gradually leak the energy in the line out into free space. This is a continuous leakage along the length of the transmission line, as contrasted to the discrete coupling out of slotted waveguide arrays. With such continuous leakage, the direction in which the energy radiates is determined by the phase velocity along the transmission line. The radiated beam will scan with frequency if the phase velocity of the transmission line changes with frequency. With independent and precise control of both the phase velocity and the rate of leakage, it is possible to curve the transmission line to fit a prescribed contour, and still enable all the leakage energy to add up in a given direction. In addition, these antennas may be so mounted on a vehicle that the leaky surface is flush with the skin of the vehicle.

Very few, if any, leaky-wave antennas have found practical application in the past, principally because it is necessary to know the phase velocity and the rate of leakage along the structure very accurately in order to design large antennas with prescribed patterns, and adequate design data have not been available either theoretically or experimentally.

This paper describes a leaky-wave antenna consisting of an inductive sheet spaced over a continuous conducting sheet, with a wave polarized in the plane of the antenna propagating between the two sheets. A sketch of this type of antenna is shown in Fig. 1. This structure is quite simple to fabricate, and the theoretical prediction of its properties is, for all practical purposes, exact. The analytical procedure utilizes a transverse-resonance technique that has been shown to be useful for a variety of radiating structures.[1–6] It is shown that this technique yields extremely precise results, enabling the phase velocity and the rate of leakage along the antenna to be predicted with an unusually high degree of accuracy.

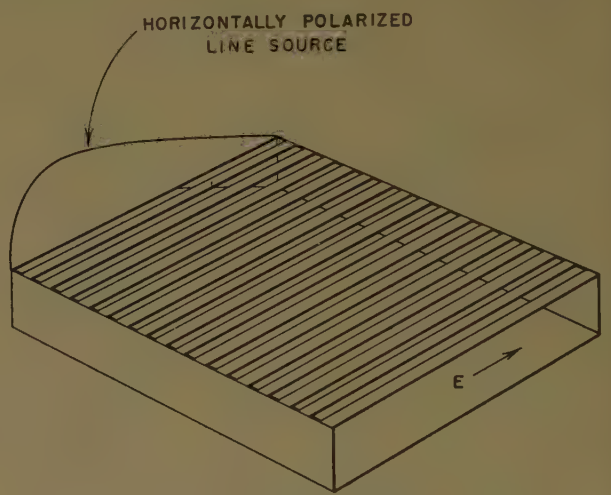


Fig. 1—Inductive sheet antenna.

The first part of this paper describes the experimental results obtained with this antenna, and compares these results with the theoretical predictions. The second part briefly outlines the analytical technique, and presents numerical results useful for designing this type of antenna.

* Manuscript received by the PGAP, March 20, 1959. The work reported in this paper was done at Stanford Research Institute on Contract AF 19(604)-1571 from the AF Cambridge Research Center.

† Stanford Research Inst., Menlo Park, Calif.

[1] R. C. Honey, "Horizontally-Polarized Long-Slot Array," Stanford Res. Inst., Menlo Park, Calif., Tech. Rept. No. 47, Contract AF 19(604)-266; August, 1954.

[2] R. C. Honey, "A Flush-Mounted Horizontally-Polarized Directional Antenna," Stanford Res. Inst., Menlo Park, Calif., Tech. Rept. 54, Contract AF 19(604)-1296; January, 1956.

[3] N. Marcuvitz, "On field representations in terms of leaky modes or eigenmodes," IRE Trans. on Antennas and Propagation, vol. AP-4, pp. 192–194; July, 1956.

[4] L. O. Goldstone and A. A. Oliner, "Leaky Wave Antennas I: Rectangular Waveguides," Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., Res. Rept., No. R-606-57, Contract AF 19(604)-2031; August, 1957. (ASTIA No. AD 146810.)

[5] L. O. Goldstone and A. A. Oliner, "Leaky Wave Antennas II: Circular Waveguides," Microwave Res. Inst., Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., Res. Rept. No. R-629–57, PIB-557, Contract AF 19(604)-2031; January, 1958 (ASTIA No. AD 152367.)

[6] R. C. Honey, "A Flush-Mounted Leaky-Wave Antenna with Predictable Patterns," Stanford Res. Inst., Menlo Park, Calif., Sci. Rept. No. 4, Contract AF 19(604)-1571; June, 1958. (ASTIA No. AD 133660.)

II. Comparison of Measured and Theoretical Performance

A. General

The experimental antenna consisted of a slightly curved conducting surface between conducting side walls that supported the inductive sheet. As shown in the sketch of Fig. 2, the aperture of the antenna was 18 inches in the E plane by 24 inches in the H plane. In the sketch, the curvature of the conducting surface is exaggerated for emphasis. Curvature was required in order to maintain the phase velocity constant along the antenna as the rate of leakage, or illumination of the aperture, was increased from zero at the beginning, then reduced to zero again at the end. The spacing between the inductive sheet and the conducting surface controlled the phase velocity in the same way that the width of conventional rectangular waveguide controls the phase velocity in the waveguide.

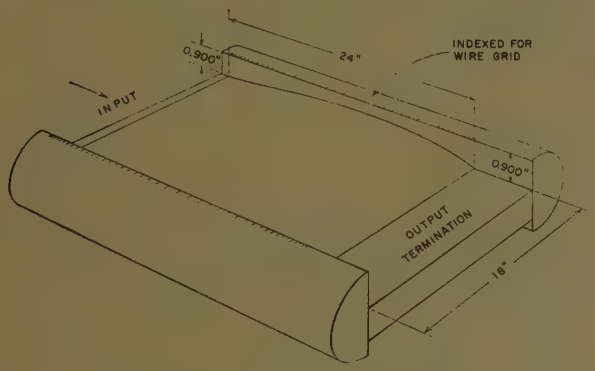


Fig. 2—Sketch of antenna geometry.

Two different inductive sheets were used for these measurements. The first conformed more closely to the theoretical model; it consisted of a large number of parallel wires stretched across the aperture. This grid was constructed by winding 0.005-inch-diameter wire completely around the structure shown in Fig. 2, varying the spacing between adjacent wires to control the amplitude of the illumination, or rate of leakage, along the aperture. In the second model the inductive sheet consisted of flat strips 0.010-inch wide printed on a thin layer of Teflon-Fiberglas laminate supported over the ground plane with polyfoam. A photograph of this latter antenna mounted on the antenna pattern range is shown in Fig. 3, complete with line source feed used for both antennas. The design of this line source is described in Section III-*C*. The important dimensions of both antennas are given in Table I in the Appendix.

An extensive series of measurements was made with both of these antennas. These measurements are described in this section and compared with the theoretical predictions.

B. Radiation Patterns

The aperture distribution which was prescribed for both antennas at the design frequency of 11.42 kmc is given by

$$f(y) = (1 - \cos y)^{1/2}\left[1 - \frac{0.85}{2\pi}(y - \sin y)\right]^{1/2}, \qquad (1)$$

where y is the normalized coordinate along the aperture, $2\pi z/L$, and L is the length of the aperture With this distribution, only $2\frac{1}{4}$ per cent of the input power remains at the end of the array for absorption in a load. This aperture distribution is shown in Fig 4. It is immediately apparent from the fact that it is an asymmetric distribution that the resultant pattern will not have deep nulls.

The H-plane gain factor corresponding to this aperture distribution is given by

$$g_H = \frac{\left|\int_0^L f(y)dy\right|^2}{\int_0^L f^2(y)dy} = \frac{0.8575^2}{0.9775} = 0.7523. \qquad (2)$$

The integration of the amplitude distribution was performed graphically, while the integration of the power distribution was done analytically.

The theoretical radiation pattern resulting from this aperture distribution at the design frequency was calculated thus: following Silver's[7] notation, the amplitude

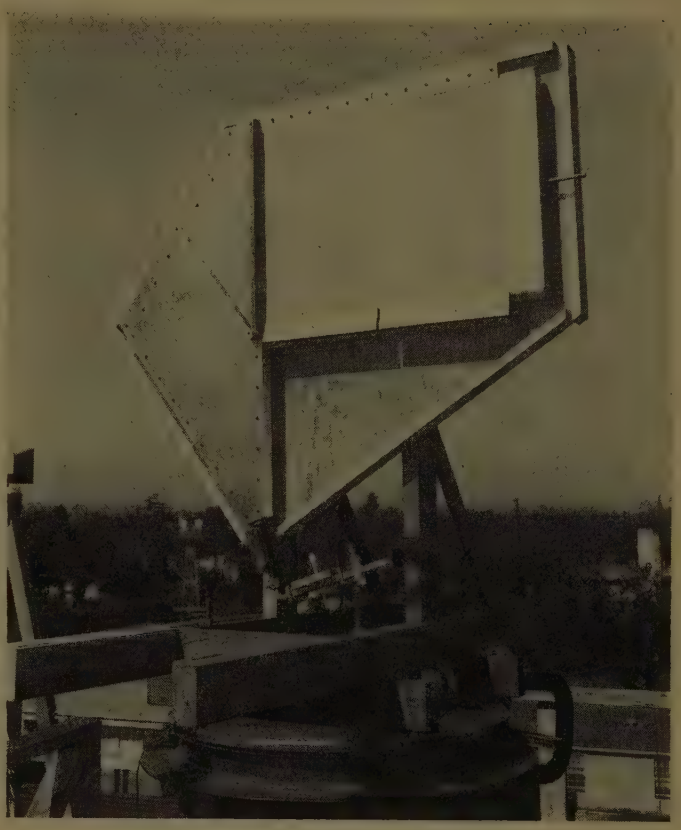


Fig. 3—Printed-strip leaky-wave antenna.

[7] S. Silver, "Microwave Antenna Theory and Design," MIT Rad, Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 12, ch. 6; 1949.

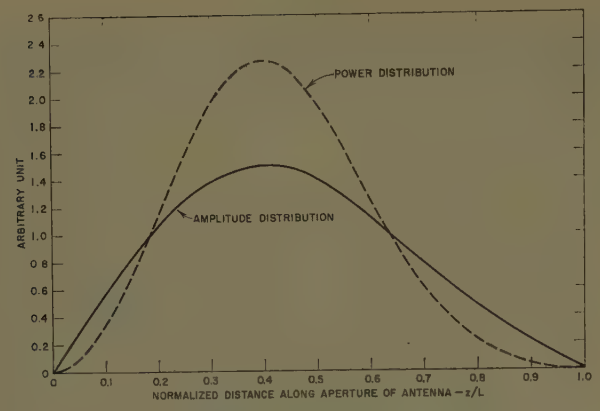


Fig. 4—Theoretical aperture distributions along array.

of the basic pattern (neglecting multiplicative constants and the element factor for the moment) is given simply by

$$g(u) = \int_0^{2\pi} f(y) e^{iuy} dy, \tag{3}$$

where u is the normalized angular variable given by

$$u = \frac{L}{\lambda} (\sin \theta - \sin \theta_0), \tag{4}$$

where λ is the wavelength and the angles are measured from the normal to the aperture. The angle θ_0 is a constant for a constant phase velocity along the antenna, and for these antennas at the design frequency of 11.42 kc,

$$\theta_0 = 54.97 \text{ degrees}. \tag{5}$$

The Appendix lists the numerical values of the various constants in these equations. Eq. (3) was integrated on a digital computer to find the normalized power pattern $p(u) = |g(u)|^2 / |g(0)|^2$ in terms of u. The values of θ satisfying (4) were then found as a function of u, and the power $p(u)$ at each point multiplied by the square of the element factor, $\cos \theta$, for the H-plane pattern of an electric dipole lying in and polarized parallel to an infinite conducting plane. The final pattern, renormalized to unity at the peak of the beam, is shown in Fig. 5(a).

The measured H-plane radiation pattern of the wire-grid antenna at the design frequency is compared with this theoretical pattern in Fig. 5(b). It can be seen that the agreement between the two is quite remarkable—the difference between them amounting to only a fraction of a decibel, on the average, even in the region 30-to-40 db below the peak of the main beam.

H-plane radiation patterns were measured for both antennas at a number of other frequencies between 7 and 13 kmc. A typical series of patterns for the wire-grid antenna is shown in Fig. 6. The angles shown on the abscissa are measured from the direction normal to the antenna, positive angles toward the end-fire direction. A similar series of patterns for the printed-strip antenna is shown in Fig. 7. It can be seen that both antennas maintain very clean, low-sidelobe H-plane patterns over very wide bands, although for the printed-strip antenna, the errors in printing the strips and supporting them accurately over the conducting sheet increased the wide-angle sidelobes to the −30-db level (the order of unity gain). The sidelobes of the wire-

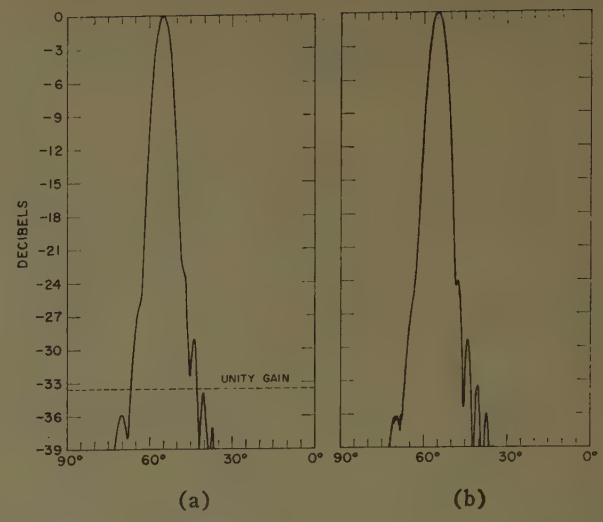


Fig. 5—Comparison of experimental and theoretical H-plane radiation patterns at 11.42 kmc. (a) Theoretical. (b) Experimental.

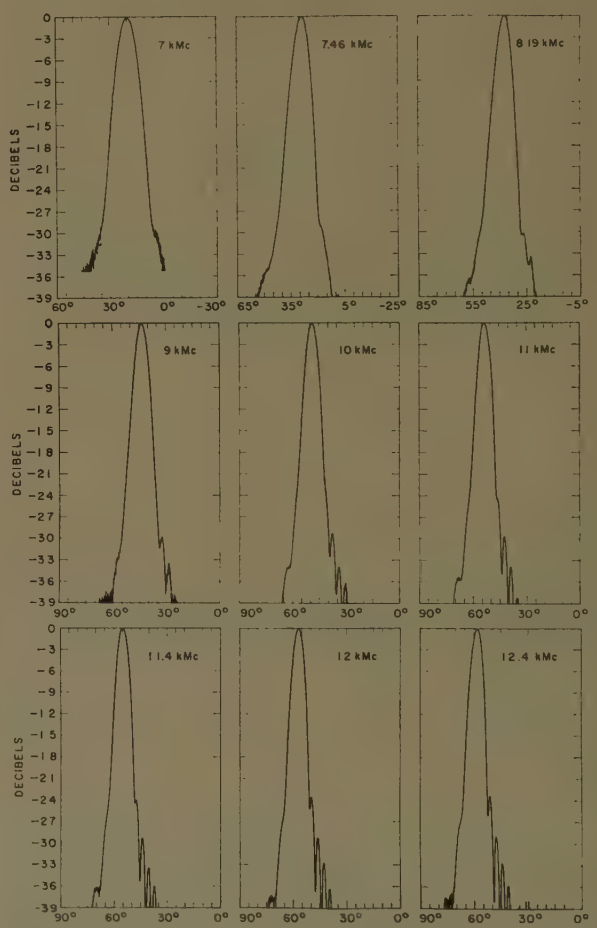


Fig. 6—H-plane radiation patterns for wire-grid antenna.

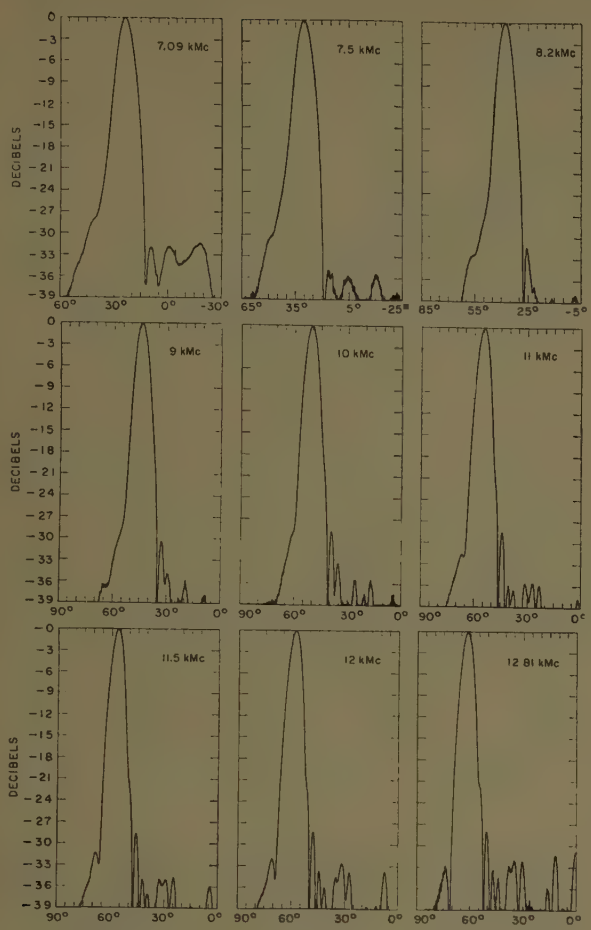


Fig. 7—H-plane radiation patterns for printed-strip antenna.

grid antenna beyond 3 or 4 beamwidths on either side of the main lobe are everywhere greater than 40 db below the main lobe at all frequencies tested.

As seen from the radiation patterns, the beam from the antenna scans in the H plane with frequency. This scanning action is caused by the change in phase velocity along the antenna as the frequency changes. The angle at which the radiation leaves the antenna is very nearly identical to the angle that the interfering TEM waves make with the side walls of a waveguide in the simple description of waveguide propagation.[8] This angle is given by

$$\sin \theta_0 = \frac{\lambda}{\lambda_g}, \tag{6}$$

where θ_0 is measured from the normal to the antenna. A very slight correction caused by the element factor, $\cos \theta$, gives the following result for the direction of the main lobe of the antenna:

$$\theta' = \theta_0 - 0.219 \tan \theta_0°. \tag{7}$$

Both theoretical beam directions, θ' and θ_0, are plotted in Fig. 8 along with the experimentally-measured beam directions for the two antennas.

[8] *Cf.* S. Ramo and J. R. Whinnery, "Fields and Waves in Modern Radio," J. Wiley & Sons, Inc., New York, N. Y., p. 373; 1953.

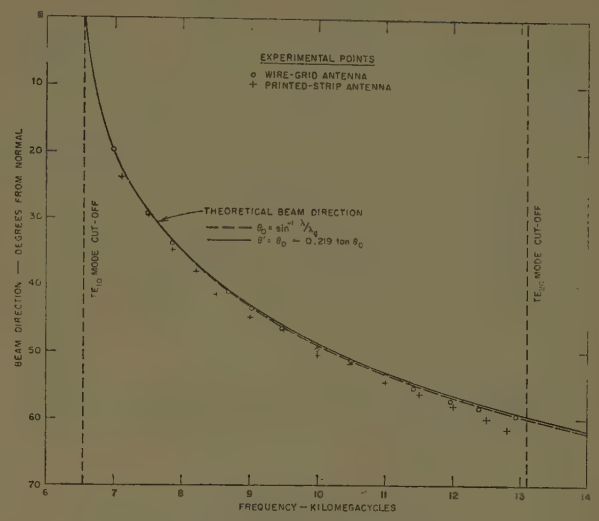


Fig. 8—Scan angle vs frequency.

The experimental values of the 3-db beamwidth as a function of frequency are plotted in Fig. 9 for both antennas. The fact that the beamwidth remains quite constant over a large part of the band is a consequence of (6) and the fact that the projected aperture L_p equals

$$L_p = L \cos \theta_0 = L\left[1 - \left(\frac{\lambda}{\lambda_g}\right)^2\right]^{1/2}, \tag{8}$$

and

$$\left(\frac{\lambda}{\lambda_g}\right)^2 = 1 - \left(\frac{\lambda}{\lambda_c}\right)^2, \tag{9}$$

where λ_c is the cutoff wavelength. Therefore, the projected aperture expressed in wavelengths becomes

$$\frac{L_p}{\lambda} = \frac{L}{\lambda_c}, \tag{10}$$

which is a constant independent of wavelength. It may be concluded that if the amplitude distribution along the aperture could be kept constant and the phase distribution linear with changes in frequency, the H-plane beamwidth from the antenna would remain independent of frequency. This conclusion is slightly qualified by the element factor, but for most practical cases this effect is negligible. The theoretical beamwidths assuming constant illumination are also shown in Fig. 9, the dashed line neglecting the element factor, the solid line taking it into account.

It can be observed from Fig. 9 that the beamwidth remains quite constant on both sides of the design frequency, but increases at the low-frequency end of the band. This increase is caused by the fact that the amplitude distribution has changed considerably from the design distribution. This change in distribution results from the increase in the rate of leakage out of all portions of the antenna as the frequency decreases.

The E-plane radiation patterns from the antenna are essentially identical to the E-plane patterns of the line source used to feed the antenna. The experimentally

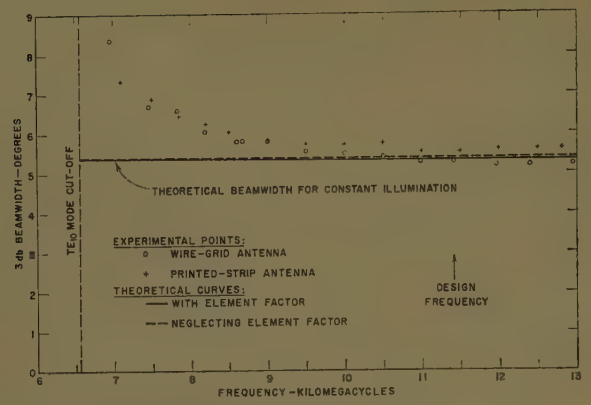


Fig. 9—H-plane beamwidth vs frequency.

measured values of the E-plane beamwidths for the complete antenna are compared in Fig. 10 to the beamwidths of the line source alone and to the theoretical beamwidth for a perfect line source of this type.

C. Gain

The gain of both antennas was carefully measured by comparing the gain with that of a standard horn. The results of these measurements are plotted in Fig. 11, along with two directivity curves calculated in different ways. The solid curve represents the directivity calculated from the measured values of the 3-db beamwidths Θ_E and Θ_H in the E and H planes, respectively, by

$$G = \frac{3.545 \times 10^4}{\Theta_E \Theta_H}, \tag{11}$$

where the beamwidths are expressed in degrees. The numerical factor is calculated from the known aperture sizes and distributions in the two planes at the design frequency and assumes that these distributions remain constant but that the effective aperture sizes change with frequency. This, of course, is not strictly true, but it provides a convenient way of approximately correlating the gain to the measured beamwidths in order to detect any unsuspected losses in the antenna.

The dashed curve in Fig. 11 represents the theoretical directivity of the antenna, assuming that the amplitude distributions remain independent of frequency and the phase distributions are linear in both planes. That is,

$$G = 4\pi \left(\frac{L_p}{\lambda}\right)\left(\frac{W}{\lambda}\right) \mathcal{G}_E \mathcal{G}_H, \tag{12}$$

where

E-plane gain factor, $\mathcal{G}_E = 0.9954$,
H-plane gain factor, $\mathcal{G}_H = 0.7523$ (2),
Projected aperture, $L_{p/\lambda} = 13\frac{1}{3}$ wavelengths (10),
Width of antenna, $W = 18$ inches.

This curve departs from the measured gain of the antennas at the low end of the band for two reasons:

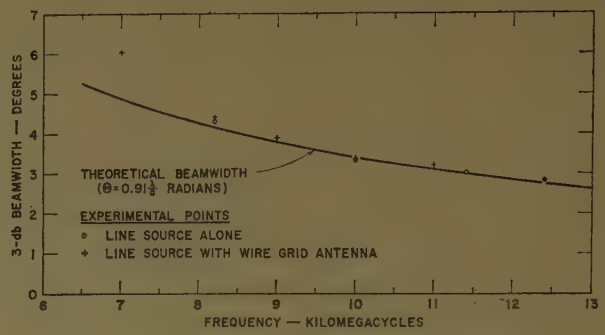


Fig. 10—E-plane beamwidth vs frequency.

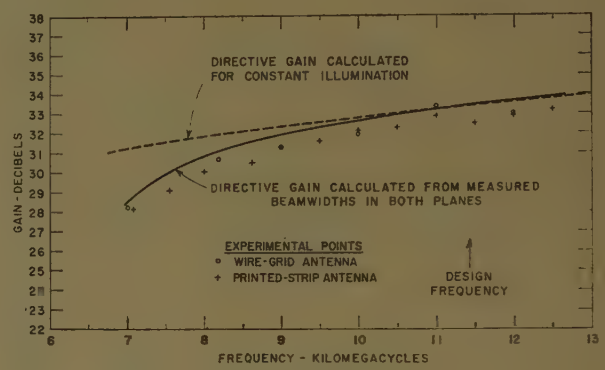


Fig. 11—Gain vs frequency for inductive-sheet antennas.

1) The phase errors in the E plane caused by finite tolerances on the spacing of the conducting sheets in the line source caused a beam broadening and a consequent loss of gain in the E plane. (See Section III-*C* and Fig. 10.)
2) The amplitude distribution along the array becomes very peaked towards the feed end of the antenna, with a consequent increase in beamwidth and a loss of gain in the H plane. (See Section II-*B* and Fig. 9.)

The general agreement, however, between the measured gain and the predicted directivity indicates that there are no unknown sources of loss within the antenna of any appreciable magnitude.

D. Input Impedance

The input impedance of the complete antenna is determined primarily by the input impedance of the line source used to feed the antenna. With the line source used here, the input VSWR remained less than 1.1 over most of the band. Fig. 12 plots the input VSWR vs frequency for the wire-grid antenna plus line source from 7-to-13 kmc.

II. Theoretical Analysis and Electrical Design

A. Analysis of Leaky-Wave Antenna

In this section, the analysis predicting the phase velocities and the attenuation constants of the various modes which can be propagated by this leaky-wave

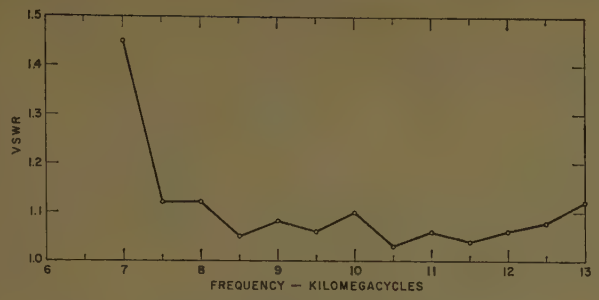


Fig. 12—Voltage standing wave ratio vs frequency for wire-grid antenna.

structure is outlined, and some numerical design data are presented.

The analysis assumes first that the antenna is an infinitely large, flat, uniform structure. The structure can then be regarded as a transmission system supporting transverse electric, or TE, waves. The complex propagation constants of these TE waves can be found by applying the transverse resonance condition, which states that the sum of the admittances or impedances looking in both directions from any plane parallel to the antenna is equal to zero. If the reference plane is taken at the plane of the inductive sheet, the equivalent circuit shown in Fig. 13(a) results, where the x coordinate is perpendicular to the plane of the antenna and a is the spacing of the inductive sheet from the conducting ground plane. Note that the normalized susceptance, B/Y_0, as well as the propagation constant perpendicular to the plane of the antenna, γ_x, may be complex if necessary.

Fig. 13(b) indicates an alternative though an entirely equivalent way of visualizing the problem. If infinitely large, perfectly conducting planes are erected perpendicular to the electric field, none of the boundary conditions are changed. If two such planes are spaced less than a half-wavelength from each other, it can then be seen that the problem of the infinitely extending array radiating into half-space is identical to the problem of a waveguide with one side wall replaced by an inductive grating radiating into an infinite parallel-plane region. The equivalent network for this problem is, of course, the same as that obtained before and shown in Fig. 13(a).

Setting the sum of the admittances in this equivalent circuit equal to zero gives

$$j\frac{B}{Y_0} = -(1 + \coth \gamma_x a). \tag{13}$$

A sheet of thin, flat strips in a homogeneous dielectric with the E field parallel to the strips is inductive so long as the spacing between strip centers, s, is small in terms of the wavelength, λ. The value of this normalized inductance is given by Marcuvitz as[9]

[9] N. Marcuvitz, "Waveguide Handbook," McGraw-Hill Book Co., Inc., New York, N. Y., pp. 285–288; 1959.

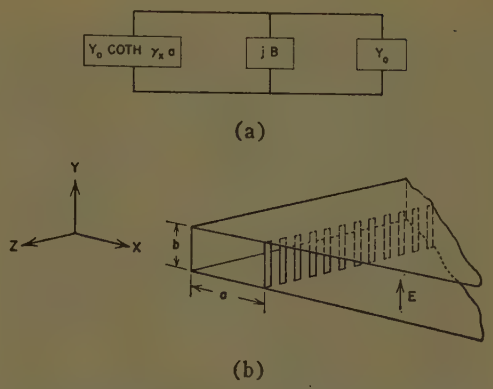


Fig. 13—Inductive-grid array. (a) Equivalent network in the plane of the grid. (b) Section of the array.

$$\frac{X}{Z_0} = \frac{s\cos\theta}{\lambda} \ln \csc \frac{\pi d}{2s},$$

where θ is the angle of incidence measured from the normal to the plane of the sheet, and d is the strip width. Thus the normalized admittance is found from

$$j\frac{B}{Y_0} = \frac{Z_0}{jX} = \left(\frac{\lambda}{j2\pi\cos\theta}\right)\frac{2\pi}{s \ln \csc \dfrac{\pi d}{2s}} \rightarrow \frac{1}{\gamma_x}\frac{2\pi}{s \ln \csc \dfrac{\pi d}{2s}}, \tag{14}$$

where the arrow indicates that the quantities comprising the imaginary propagation constant in the equation have been replaced by the more general complex propagation constant appropriate to this problem.

Combining (13) and (14) gives

$$1 + \coth \gamma_x a = -\frac{C}{\gamma_x a}, \tag{15}$$

where

$$C = \frac{2\pi a}{s \ln \csc \dfrac{\pi d}{2s}}, \tag{16}$$

a purely real function of the geometry. If the spacing is not sufficiently small in terms of wavelength, correction terms must be added to the simple expression above. Thus, (16) must be replaced by

$$\frac{C}{a} = \frac{2\pi}{s\left[\ln\left(\csc\dfrac{\pi d}{2s}\right) + F\right]}, \tag{17}$$

where the correction term, F, from MacFarlane,[10] is shown in Fig. 14 as a function of angle of incidence, θ_0, and spacing, s. The values of the spacings, s, found from

[10] G. G. MacFarlane, "Surface impedance of an infinite parallel-wire grid at oblique angles of incidence," *J. IEE (London)*, vol. 93, pt. III-A, no. 10, pp. 1523–1527; 1946.

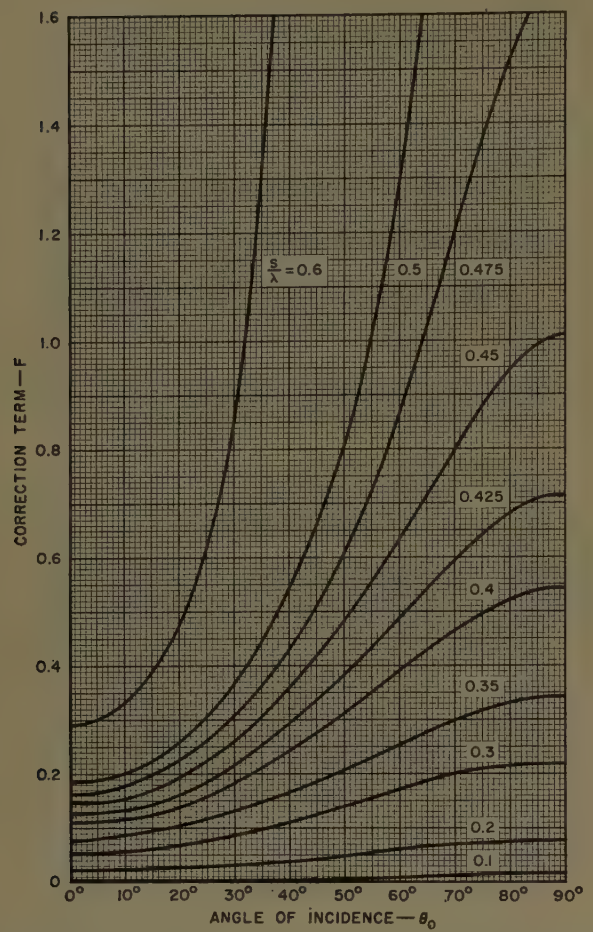


Fig. 14—Correction term F as a function of angle of incidence for various values of the spacing.

this formula and used in the present antenna for $d=0.010$ inch are tabulated in the Appendix. These results apply equally well to a grid of round wires by simply letting d represent twice the wire diameter instead of the strip width, so long as the strip width is a small fraction of a wavelength. Wait[11] has derived a more exact value for the correction term, F, taking into account the presence of the conducting plane. His results differ appreciably from those of MacFarlane only for much larger wire spacings than used in the present antenna.

The quantities of interest, however, are the real and imaginary parts of the propagation constant in the z direction, or along the array. These can be found from γ_x since the wave equation requires that

$$\gamma_x^2 + \gamma_z^2 + k^2 = 0, \tag{18}$$

for no variations in the y direction ($\gamma_y=0$), where, in free space,

$$k^2 = \omega^2\mu\epsilon = \left(\frac{\omega}{c}\right)^2 = \left(\frac{2\pi}{\lambda}\right)^2,$$

c = velocity of light,

$\omega = 2\pi f$, the angular frequency.

Separating the propagation constants into real and imaginary parts,

$$\gamma_x a = \xi_1 + j\xi_2,$$

$$\gamma_z = \alpha + j\beta, \tag{19}$$

it is possible to find solutions for α and β in terms of ξ_1, ξ_2 and k in two regions:

In the "propagating" region $(\xi_2^2-\xi_1^2)<(ka)^2$:

$$\alpha a = \sqrt{(ka)^2 - (\xi_2^2 - \xi_1^2)}\sqrt{\frac{(1+x^2)^{1/2}-1}{2}}, \tag{20}$$

$$\beta a = \sqrt{(ka)^2 - (\xi_2^2 - \xi_1^2)}\sqrt{\frac{(1+x^2)^{1/2}+1}{2}}, \tag{21}$$

where

$$x^2 = \frac{4\xi_1^2\xi_2^2}{[\xi_2^2 - \xi_1^2 - (ka)^2]^2}.$$

In the "cutoff" region $(\xi_2^2-\xi_1^2)>(ka)^2$

$$\alpha a = \sqrt{(\xi_2^2 - \xi_1^2) - (ka)^2}\sqrt{\frac{(1+x^2)^{1/2}+1}{2}}, \tag{22}$$

$$\beta a = \sqrt{(\xi_2^2 - \xi_1^2) - (ka)^2}\sqrt{\frac{(1+x^2)^{1/2}-1}{2}}. \tag{23}$$

The guide wavelength in the z direction, λ_g, can be found from βa with the simple relationship,

$$\frac{\lambda}{\lambda_g} = \frac{\lambda}{2\pi a}(\beta a). \tag{24}$$

In order to illustrate the type of behavior which occurs in this inductive sheet structure, αa and λ/λ_g have been plotted vs $\lambda/2a$ for several values of the parameter C in Figs. 15 and 16. The same factors for the TE_{10} and TE_{20} waveguide modes are also plotted to indicate the way in which the modes in the inductive wall guide differ from the modes in standard rectangular waveguide. The "propagating" and "cutoff" regions referred to above are defined simply by extrapolation from the behavior of unperturbed waveguide, even though the terms are not strictly accurate in this case. The dividing point between the two regions, or cutoff point shown by the small circles in the figures, occurs when $(ka)^2=\xi_2^2-\xi_1^2$.

[11] J. R. Wait, "Reflection from a wire grid parallel to a conducting plane," *Can. J. Phys.*, vol. 32, pp. 571–579; September, 1954.

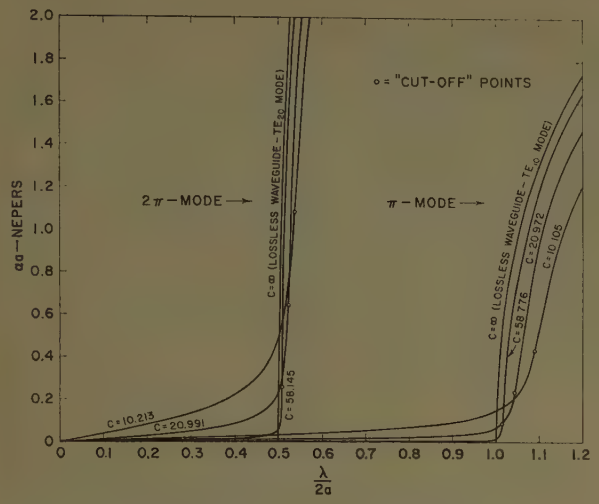


Fig. 15—The normalized attenuation constant αa vs $\lambda/2a$ for several values of the parameter C.

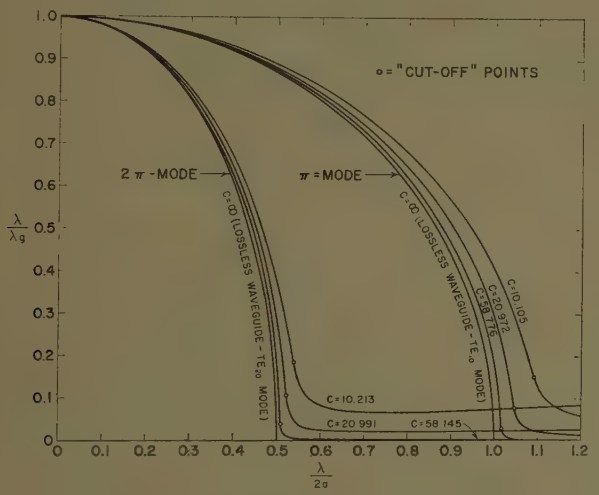


Fig. 16—The free-space wavelength divided by the guide wavelength vs $\lambda/2a$ for several values of the parameter C.

The solutions of (15) and (20)–(24) are plotted in Figs. 17 and 18, where $\lambda/2a$ and $\alpha\lambda$ are shown as functions of $C\lambda/2a$, with λ/λ_g as a parameter. These plots are particularly useful for designing antennas in which the phase velocity, or λ/λ_g, is a constant along the array since the abscissa, $C\lambda/2a$, is a dimensionless constant independent of a (since C contains a as a factor) and proportional to wavelength. In a similar way, the ordinate in Fig. 18, $\alpha\lambda$, expressed in nepers, is independent of the variable a. Thus, given the frequency and the desired angle of the radiation, or λ and λ_g, the width, a, and the attenuation constant, α, can be found as a function of the strip geometry, C/a, from Figs. 17 and 18, respectively.

B. Design of Amplitude and Phase Distribution

In order to design an antenna with the data presented in the previous section, it is first necessary to

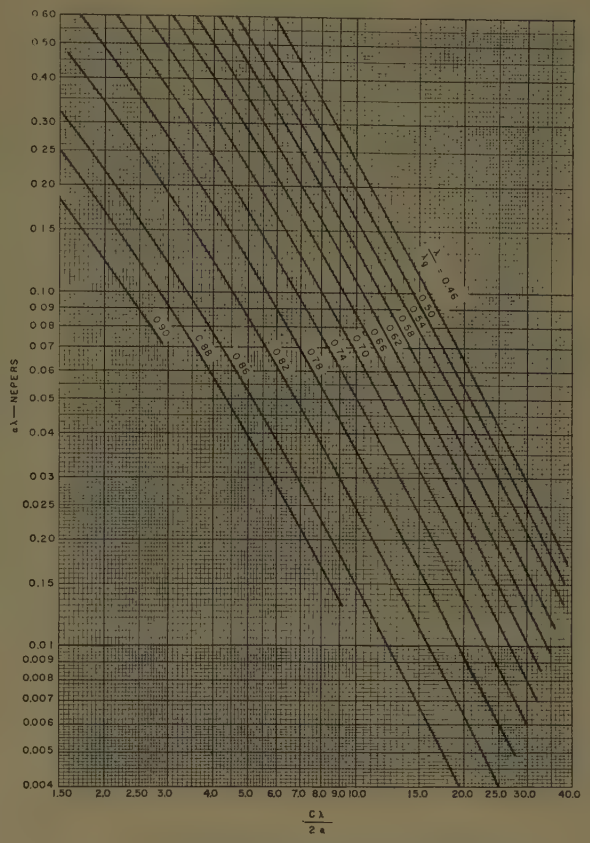


Fig. 17—A plot of $\alpha\lambda$ vs $C\lambda/2a$ for several values of the parameter λ/λ_g.

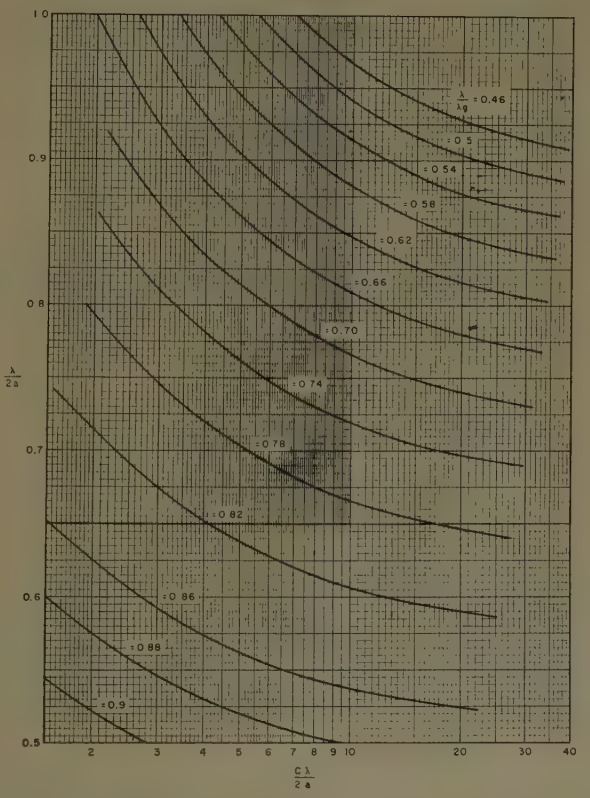


Fig. 18—A plot of $\lambda/2a$ vs $C\lambda/2a$ for several values of the parameter $\lambda/\lambda g$.

specify the phase and amplitude distributions along the aperture which will produce the desired radiated beam. It should be remembered that the smoother these distributions are, the more closely the theory will apply. Thus, if in the neighborhood of a radiating segment of the array the boundary conditions do not differ appreciably from those existing in an infinite, uniform array, then the radiation from this segment should not differ appreciably from the radiation from an equal segment in an infinite, uniform array. It is also worth noting that the larger the size of the antenna, in both the E and the H planes, the more nearly will it correspond to the infinitely large antenna treated analytically.

The amplitude distribution specified for the antenna may be related to the exponential attenuation constant, α, which is used in the theoretical derivation in the following manner. Note that the power inside the array at a point z along it, $P(z)$, is related to the power radiated per unit distance along the array, $P_r(z)$, by

$$P_r(z) = -\frac{dP(z)}{dz} = 2\alpha(z)P(z). \tag{25}$$

Note also that

$$P(z) = P(0) - \int_0^z P_r(\xi)d\xi, \tag{26}$$

where $P(0)$ is the power into the antenna at $z=0$ and ξ is simply the variable of integration along the aperture. The total power radiated by the antenna is given by

$$P(0) - P(L) = \int_0^L P_r(\xi)d\xi = FP(0), \tag{27}$$

where F is the fraction of the input power which is radiated by the antenna and L is the total length of the aperture. Eqs. (25), (26), and (27) can be solved for the attenuation constant α, giving

$$2\alpha(z) = \frac{f^2(z)}{\frac{1}{F}\int_0^L f^2(\xi)d\xi - \int_0^z f^2(\xi)d\xi}, \tag{28}$$

where $f(z)$ is the amplitude distribution across the aperture, given by $P_r(z) = f^2(z)$.

Applying (28) to the aperture distribution given by (1) gives the values for $L\alpha(z)$ shown in the Table in the Appendix.

C. Line Source Feed

In order to obtain a broad-band horizontally-polarized line source with no obstructions in the aperture, an asymmetrically fed pillbox antenna was selected. Fig. 19 shows the important dimensions of the line source, which was designed on the basis of geometric optics to provide an unobstructed aperture 18 inches in length. It is fed from standard 0.400- by 0.900-inch ID waveguide, and the height of the pillbox is maintained at 0.900 ± 0.002 inch throughout. The finite tolerances that could be maintained on this spacing limited the performance of the line source at the low end of the frequency band as the nominal cutoff frequency of 6557 mc was approached.

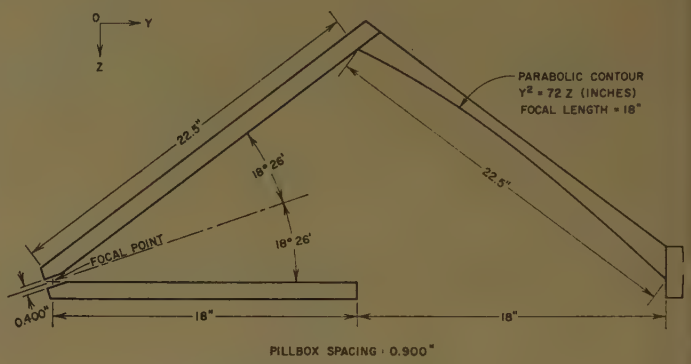


Fig. 19—Horizontally-polarized line source.

IV. Conclusion

It has been shown that the transverse-resonance technique used to predict the behavior of leaky-wave antennas gives very precise results, permitting very close control of the phase and of the amplitude distributions across the aperture of an antenna. The particular leaky-wave antenna described in this report can be flush-mounted behind the skin of a vehicle such as an airplane, and requires between one-half and one wavelength of depth if not dielectrically loaded.

With such precise and independent control of the phase and amplitude distributions over the aperture, it is possible to build antennas which radiate very accurately-shaped beams, or very low sidelobe beams. In addition, such antennas can be curved in the H plane to fit an airframe contour and the phase velocity can be correspondingly adjusted so that the antenna still radiates a pencil beam or a shaped beam in a given direction. In addition, it is a relatively simple matter to generalize the analyses to take into account various modifications such as partial or complete dielectric loading in the space between the inductive sheet and the conducting wall; this would provide better mechanical support for the inductive sheet and allow the antenna to radiate closer to the end-fire direction without TE_{20} mode problems. In a similar manner, a capacitive sheet placed midway between the inductive sheet and the conducting wall, either with a sheet of dielectric or with conducting strips or wires running perpendicular to those in the inductive sheet, would extend the fre-

quency bandwidth of the present antenna, allowing it to radiate closer to the end-fire direction.

The frequency-scanning properties of these antennas may also find application. The scan angle associated with a given frequency change may be increased by operating in a more dispersive region, or by increasing the dispersion with dielectric or periodic loading.

Appendix

Dimensions of Leaky Wave Antenna

Length of aperture (H-plane), $L = 24$ inches
Width of aperture (E-plane), $W = 18$ inches
Wire diameter, $d/2 = 0.005$ inch
Design frequency, $f = 11.42$ kmc.
Design wavelength, $\lambda = 2.624$ cm $= 1.033$ inches
Design guide wavelength, $\lambda_g = 3.205$ cm $= 1.262$ inches
At the design frequency,

$$\sin \theta_0 = \frac{\lambda}{\lambda_g} = 0.81883,$$

$$\theta_0 = 54.97°$$

TABLE I

z/L	$L\alpha(z)$ nepers	$\alpha(z)$ nepers/inch	C/a inches^{-1}	a inches	s inches	z inches
0	0	0	∞	0.900	0.005	0
0.05	0.04163	0.001735	73.8	0.886	0.0617	1.2
0.10	0.16323	0.006801	36.9	0.8732	0.0943	2.4
0.15	0.35682	0.01487	24.8	0.8607	0.122	3.6
0.20	0.61266	0.02553	18.7	0.8486	0.148	4.8
0.25	0.92113	0.03838	15.0	0.837	0.1723	6.0
0.30	1.2735	0.05306	12.55	0.8262	0.195	7.2
0.35	1.6622	0.06926	10.8	0.816	0.216	8.4
0.40	2.0790	0.08663	9.5	0.8062	0.237	9.6
0.45	2.5154	0.10481	8.5	0.7973	0.256	10.8
0.50	2.9565	0.12319	7.70	0.7892	0.274	12.0
0.55	3.3797	0.14081	7.08	0.7817	0.291	13.2
0.60	3.7460	0.15608	6.62	0.7759	0.304	14.4
0.65	3.9923	0.16635	6.36	0.772	0.314	15.6
0.70	4.0265	0.16778	6.30	0.7712	0.316	16.8
0.75	3.7409	0.15587	6.63	0.7762	0.305	18.0
0.80	3.0696	0.12790	7.55	0.7874	0.278	19.2
0.85	2.0894	0.08687	9.51	0.8063	0.237	20.4
0.875	1.5503	0.06460	11.23	0.8185	0.210	21.0
0.90	1.0441	0.04350	14.03	0.8332	0.180	21.6
0.925	0.60815	0.02534	18.8	0.849	0.147	22.2
0.95	0.27605	0.01150	28.15	0.8652	0.112	22.8
0.975	0.06972	0.002905	56.2	0.882	0.073	23.4
1.00	0	0	∞	0.900	0.005	24.0

The Unidirectional Equiangular Spiral Antenna*

JOHN D. DYSON†

Summary—**Circularly polarized unidirectional radiation, over a bandwidth which is at the discretion of the designer, is obtainable with a single antenna. The antenna is constructed by wrapping balanced equiangular spiral arms on a conical surface. The non-planar structure retains the frequency-independent qualities of the planar models, and, in addition, provides a single lobe radiation pattern off the apex of the cone. Practical antennas have been constructed with radiation patterns and input impedance essentially constant over bandwidths greater than 12 to 1 and there is no reason to assume that these cannot be readily extended to more than 20 or 30 to 1.**

* Manuscript received by the PGAP, July 27, 1958; revised manuscript received April 27, 1959. This work was supported by Wright Air Dev. Cen. under Contract No. AF 33(616)3220.

† Antenna Lab., Dept. of Elec. Eng., University of Illinois, Urbana, Ill.

Introduction

THE principle of scaling in electromagnetic theory has been used extensively for many years in the testing of antennas. More recently, it has been applied to the design of antennas that have unique and useful characteristics as the frequency of operation is varied.[1] The characteristics of these antennas are essentially frequency independent over bandwidths which, although they are not unlimited, may be easily extended to 20 to 1, or more. The general approach is to design

[1] V. H. Rumsey, "Frequency Independent Antennas," 1957 IRE National Convention Record, pt. 1, pp. 114–118.

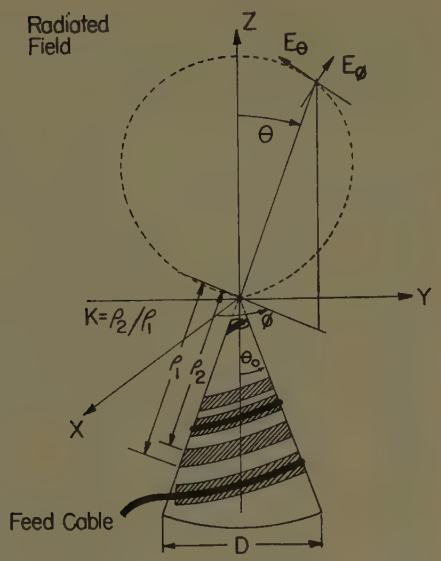


Fig. 1—Balanced conical antenna with radiation pattern coordinate system.

a radiating structure such that successive applications of a scaling factor result in structures which are identical with the original, or which, at most, differ by some rotation about an axis through the origin of the original structure. Although the structure must be infinite in size to fulfill the scaling condition of exact equivalence, it has been found that there is a class of structures that may be truncated and still retain, over an extremely wide range of frequencies, the characteristics of infinite size. This class of antennas is based upon a plane curve, the equiangular or logarithmic spiral, which possesses the property of having a change of scale merely equivalent to a rotation. Several such curves can be used to outline antennas on plane or conical surfaces. The basic characteristics of the planar antenna have been covered elsewhere;[2] the present paper is concerned with the nonplanar antenna.

It is possible to retain the desirable frequency-independent characteristics of the planar antenna and, in addition, confine the radiation to one hemisphere by forming the planar arms down onto a conical surface.[3] The nonplanar antenna so constructed, with a small cone angle, has a unidirectional rotationally-symmetric radiation pattern. The maximum radiation occurs on the antenna axis off the apex of the cone. These balanced conical spirals can be constructed to operate over bandwidths comparable to those of the basic planar structure.

[2] J. D. Dyson, "The Equiangular Spiral Antenna," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-7, pp. 181–187; April, 1959.

[3] J. D. Dyson, "The Unidirectional Equiangular Spiral Antenna," University of Illinois Antenna Lab. Tech. Rep. No. 33, Contract AF 33(616)3220, Wright Air Dev. Cen., Wright-Patterson Air Force Base, Ohio, July 10, 1958.

Fig. 2—Conical antenna, $D \simeq 20.7$ cm, $d \simeq 3$ cm, $b = 0.053$, $\alpha = 78°$, $K = 0.925$, $\theta° = 10°$.

The Nonplanar Antenna

The conical spiral antenna is defined as in Fig. 1. On the surface of revolution $\pi - \theta = \theta_0$; the edges of the arms are defined by

$$\rho_1 = e^{(a \sin \theta_0)\phi} = e^{b\phi}$$

and

$$\rho_2 = e^{b(\phi - \delta)} = K\rho_1.$$

A rotation of both curves by 180 degrees will provide the second arm of a balanced structure. The edges of the antenna, so defined, will maintain a constant angle α with the radius vector for any cone angle θ_0; however, the antenna will spiral more tightly; *i.e.*, b will decrease, as θ_0 decreases from $\pi/2$.

The antennas may be constructed by forming metal arms on a suitable conical support or may conveniently be etched on a flexible teflon-impregnated fiberglass material and then formed into a cone. Fig. 2 shows an etched antenna fed with RG 141/U cable bonded to the arms. As previously indicated for the planar antennas, these are balanced structures and a balanced feed is necessary for optimum performance. The feed may be brought in along the axis of the antenna by a balanced feed line or by an unbalanced line and balancing transformer or balun. The bandwidth of this latter method, of course, depends upon the bandwidth of the balun. The rapid decay of the current along the arms, observed in the planar case, is present in a modified form in these nonplanar antennas. This makes possible the same method of feeding the structure; *i.e.*, the cable is bonded to one antenna arm and is carried to the origin

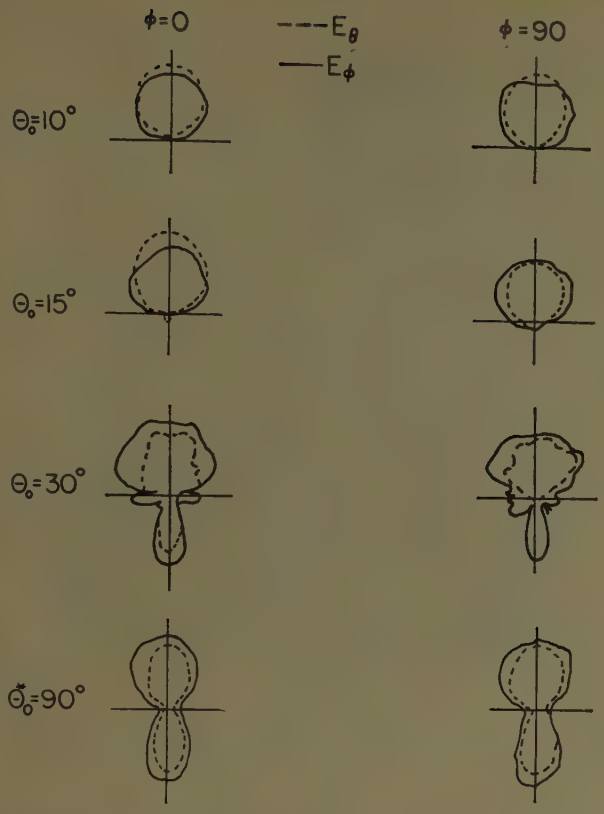


Fig. 3—Electric radiation patterns as a function of θ_0. $K=0.925$, $L=150$ cm, $\alpha=73°$, $b=(0.30 \sin \theta_0)$, $f=2000$ mc. * Dissymmetry between front, back lobes ($\theta_0=90°$) due to mount.

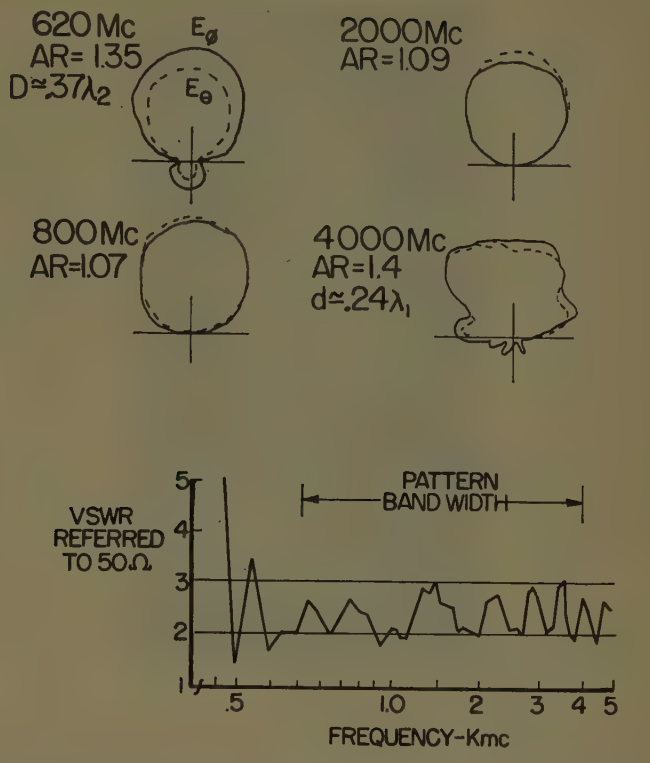


Fig. 4—Electric field patterns and VSWR of conical antenna. $\theta=10°$, $K=0.925$, $D=17.7$ cm, $\alpha=73°$, $b=0.053$, $\phi=0°$, AR=axial ratio on axis.

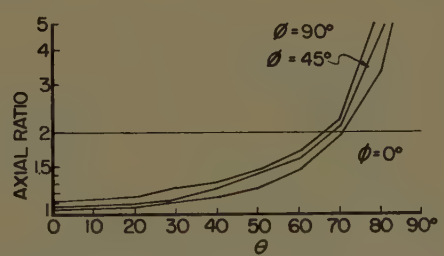


Fig. 5—Polarization of the e field of typical conical antenna as a function of angle off axis.

where the center conductor is tied to the opposite arm. A dummy cable on the opposite arm will maintain symmetry. Since the ends of the antenna arms do not carry appreciable antenna currents except at the very lowest frequency of operation, the arms themselves act as an "infinite balun," the feed terminals are isolated from ground in a balanced manner and the outside of the feed cable beyond the antenna arms does not carry a significant amount of antenna current. However, as the frequency of operation is decreased, a point will be reached where these conditions are violated; this should be considered below the range in which the antenna may be expected to operate satisfactorily.

Operation in Free Space

The change in the radiation pattern of a balanced antenna as it is depressed from a plane onto the surface of a cone as a function of the cone angle, θ_0, is indicated in Fig. 3 above. There is a marked increase in the front-to-back ratio for included cone angles of 30° (θ_0 of 15°) or less. The antenna radiates off the apex of the cone. There is no basic tilt to the pattern, and for small b (order of 0.05) the lobe is rotationally symmetric. The pattern rotates with frequency (as previously observed in the planar case), but this rotation is masked by the pattern symmetry. For $\theta_0=10°$ and $\alpha=73°$, the half-power beamwidths are typically 70° for E_θ polarization and 90° for E_ϕ polarization.

Typical radiation patterns are shown in Fig. 4. As is true with the planar antenna, the pattern bandwidth is at the control of the designer. The upper frequency, f_1, depends upon the diameter, d, of the truncated apex and the lowest useable frequncy, f_2, upon the base diameter, D. For the smaller sizes of feed cable (RG 141/U or smaller), $\theta_0=10°$ and parameters of the order of $K=0.85$ to 0.9, and $\alpha=73°$, an apex diameter of $\lambda_1/4$ and a base diameter of $3\lambda_2/8$ should provide a front-to-back ratio of 15 db, or greater, and an axial ratio well below 2 on axis. It is worth noting that the radiated field is circularly polarized well off axis, as indicated in Fig. 5.

The method of feeding the conical antenna, by embedding the cable on the antenna arms, leads to a very interesting, and what may well be the most useful

(a) (b)

Fig. 6—(a) Wire version of conical antenna, $D \simeq 30$ cm, $d \simeq 5.5$ cm, $\alpha = 73°$, $\theta = 10$, RG 8/U cable. (b) Wire version of a conical antenna, supported by polystyrene ribs. $D \simeq 30$ cm, $d \simeq 8.5$ cm, $\alpha = 73°$, $\theta_0 = 10°$, RG 8/U cable.

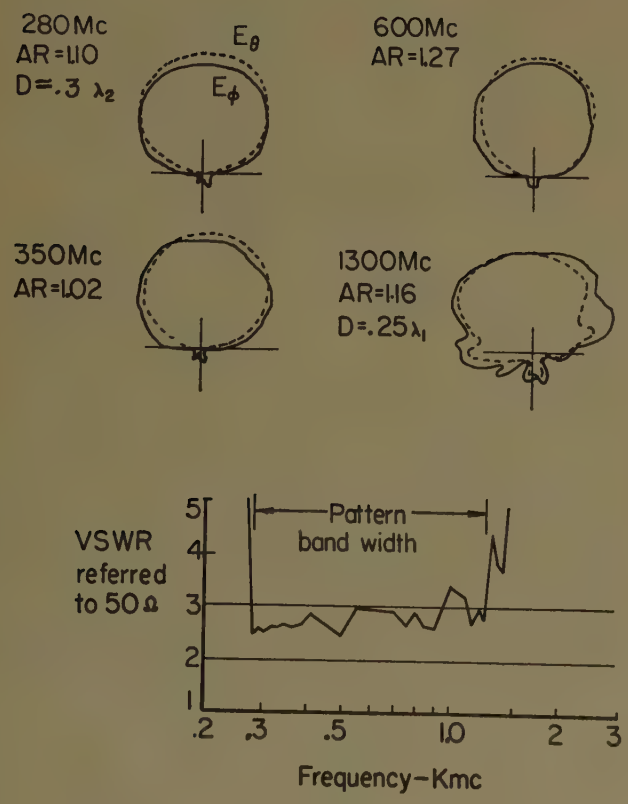


Fig. 7—Electric field radiation patterns and VSWR of wire antenna [Fig. 6 (a)]. $\theta = 10°$, $b = 0.053$, $D = 30$ cm, $\alpha = 73°$, $\phi = 0°$, K = not constant, cable = RG 8/U, AR = axial ratio on axis.

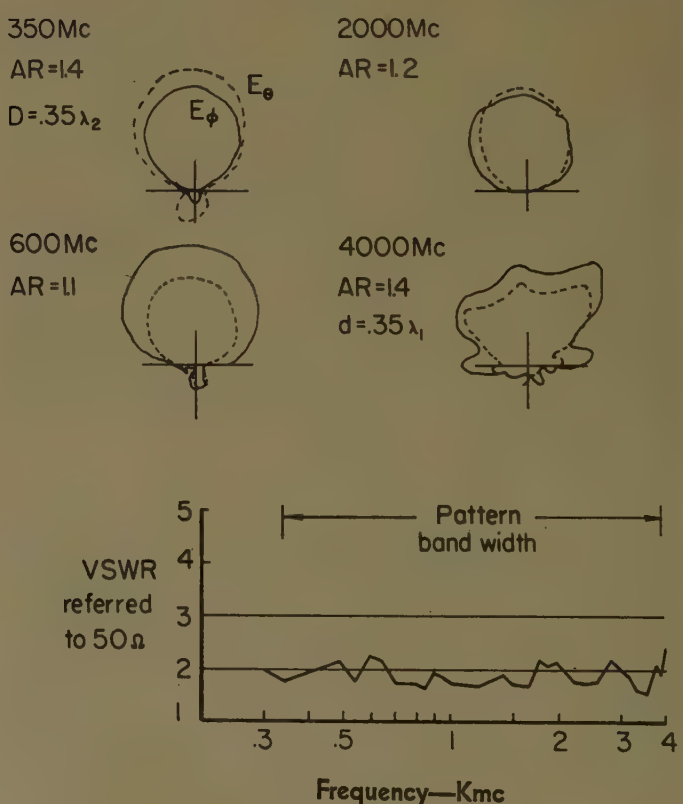


Fig. 8—Electric field radiation patterns and VSWR of conical antenna with RG 141/U arms. $\theta = 10°$, $b = 0.053$, $D = 30$ cm, $\alpha = 73°$, $\phi = 0°$, K = not constant, AR = axial ratio on axis.

version of the antenna. If the expanding arms are decreased to a narrow width and allowed to degenerate into constant width structures, the arms may be formed by the cable alone, as in Figs. 6(a) and 6(b). In this wire version, the feed cable, RG 8/U, becomes one arm. At the apex, the center conductor is carried over and bonded to the outer braid of a dummy cable which forms the opposite arm of a balanced antenna. Radiation pattern characteristics are indicated in Fig. 7. For this particular antenna, patterns were excellent down to a base diameter of approximately $0.3\lambda_2$, and, as for the true equiangular antennas referred to previously, up to an apex diameter of approximately $\lambda_1/4$. An apex diameter of approximately four centimeters appears to be a practical limit for cable of this size, which indicates an upper useful frequency of around 1600 or 1700 mc. However, the bandwidth can, as before, be extended downward to any desired frequency by a suitable extension of the cone.

Since the pattern range at this laboratory has a practical limit from 0.3 to 12 kmc, it is convenient to show the bandwidth potential by going up in frequency. Fig. 8 shows the patterns over an 11 to 1 range of frequencies for an antenna identical except for the fact that it was constructed of RG 141/U cable. The apex could then be carried to a diameter of approximately two centimeters. As indicated in Figs. 7 and 8, for a given antenna size, those structures with wider arms (large cable in the wire version, or smaller K in the true equiangular version) may be operated down to a slightly longer wavelength.

Antennas such as the one in Fig. 6 could most conveniently be constructed of solid wall $\frac{3}{8}$ inch coaxial cable and should be most useful structures in the VHF and UHF regions for telemetering and similar purposes.

The input impedance of these antennas appears to be well behaved over at least as wide a bandwidth as the radiation pattern. The mean impedance level appears to slowly decrease with decreasing cone angle. The variation in the measured impedance of the antennas referred to in Fig. 3, over the frequency range from 0.5 to 2.5 kmc, is indicated in Table I.

TABLE I

MEASURED INPUT IMPEDANCE OF BALANCED CONICAL ANTENNAS AS A FUNCTION OF θ_0 ($K=0.925$, $L=150$ cm, $b=(0.303 \sin \theta_0)$, $\alpha=73°$)

θ_0	Approximate Mean Impedance	Maximum SWR Referred to Mean
10°	129	1.9 :1
15°	147	1.9 :1
30°	153	1.95:1
90°	164	2.1 :1

The impedance is influenced by the construction at the terminal region. Since these antennas were fed with a coaxial cable of 0.140-inch diameter, this cable dominates the terminal region and undoubtedly contributes to the mean impedance level. However, this appears to be the only practical method of feeding these structures if the full bandwidth potentiality is to be realized. Hence, the input impedance with a practical size feed cable bonded to the arms is of interest.[4]

As indicated in Figs. 4, 7 and 8, these antennas would be a better match to a 100-ohm cable than to the 50-ohm cable used. However, the stability of the VSWR with frequency, over the pattern bandwidth, is apparent.

Operation over a Ground Plane

A previous investigation of the unbalanced conical antenna fed against a ground plane with the apex of the cone on the ground plane had demonstrated the pattern rotation and some of the characteristics of this version.[5] However, the radiation pattern of the unbalanced antenna was tilted off axis, and the pattern rotation caused the antenna to be of limited usefulness.

The balanced antenna referred to in Fig. 4 was placed with its base on a 12-foot square ground plane and the arms bonded to the ground plane. The presence of the ground plane narrows the beamwidth somewhat at the lower frequencies but does not deteriorate the radiation pattern, and, above 600 mc, the VSWR of the antenna was essentially unaffected by the ground plane.

Operation in a Flush-Mounted Cavity

A very limited investigation indicates that it is possible to place this antenna in a conical cavity and obtain essentially unaffected operation over the lowest portion of the bandwidth of the antenna. However, the cavities investigated have imposed on over-all frequency bandwidth of about 4 to 1.[6]

Increasing the Beamwidth

The pattern beamwidth appears to be related to the rate of spiral and varies from around 70° for an angle α of 74°, to approximately 180 or 190° for an α of 45°. Preliminary patterns for an antenna with this latter angle are shown in Fig. 9. These are of some interest because they are within approximately 6 db of being those of a circularly polarized isotropic source in one hemisphere.

[4] Since preparation of this paper a new balun has been proposed that should find wide use with these structures. See R. H. DuHamel and F. R. Ore, "Log periodic feeds for lens and reflectors," 1959 IRE National Convention Record, vol. 7, pt. 1, pp. 128–138.

[5] R. L. Carrel, "Experimental Investigation of the Conical Spiral Antenna," University of Illinois Antenna Lab. Tech. Rep. No. 22, Contract AF 33(616)-3220, University of Illinois, Urbana, May, 1957.

[6] J. D. Dyson, "The Non-Planar Equiangular Spiral Antenna," *Proc. Eighth Annual Symp. on the USAF Antenna Res. and Dev. Program*, Robert Allerton Park, Monticello, Ill., October, 1958.

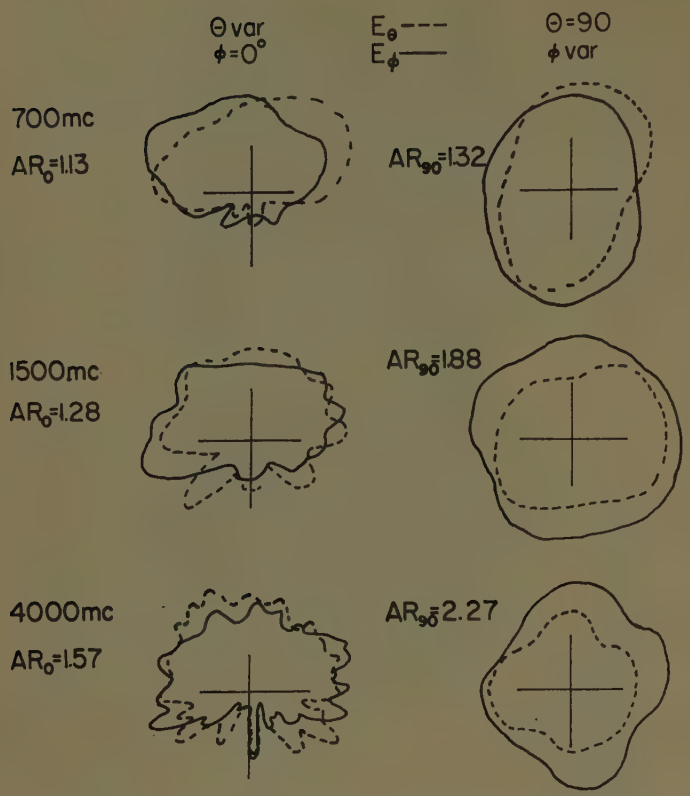


Fig. 9—Radiation patterns of balanced conical antenna with $L=78.3$ cm, $D=19.5$ cm, $H=54.5$ cm, $b=0.174$, $K=0.75$, $\theta_0=10°$, $\alpha=45°$, AR_0=axial ratio at $\theta=0°$, AR_{90}=axial ratio at $\theta=90°$.

Conclusion

An investigation of the balanced equiangular spiral antenna formed on a conical surface has shown that unidirectional, circularly-polarized single-lobe radiation patterns may be obtained over bandwidths that are at the discretion of the designer. The input impedance remains relatively constant over this pattern bandwidth.

Antennas of practical form may be constructed by shaping metal arms on a conical surface or by printed circuit techniques. A simple and most useful form is the wire version where the antenna arms consist of coaxial cable only. Such structures should find wide use at the VHF and UHF frequencies.

Operation of these nonplanar antennas is possible in free space, or with the antenna base over a ground plane. They need not be isolated from the ground plane. Flush mounted operation is possible with some sacrifice in bandwidth.

Acknowledgment

The author is pleased to acknowledge the helpful discussions he had with Dr. P. E. Mayes and the assistance he received from R. L. Jones, P. W. Arnold and other members of the University of Illinois Antenna Laboratory who contributed to the measurements

Closely-Spaced Transverse Slots in Rectangular Waveguide*

RICHARD F. HYNEMAN†

Summary—**The traveling-wave modes associated with an infinite, periodic structure are considered. An approximate equation for the propagation constants of these modes is derived through the use of Fourier analysis and an approximate application of the reaction concept. In the homogeneous case considered, it is found that two dominant modes may exist: an attenuated fundamental mode representing a perturbation of the dominant mode of a closed rectangular waveguide, and an unattenuated surface wave, which is similar to the wave associated with a corrugated surface waveguide. By means of the appropriate variation of physical parameters, including the slot length and spacing, essentially independent control of the attenuation constant and phase velocity of the fundamental mode is possible over a wide range. Typical curves of the propagation constant in terms of these parameters are given, and the results of experimental measurements are shown to be in close agreement with the theory.**

Introduction

In the field of microwave antennas, various traveling-wave devices such as the Channel Guide[1] and the Traveling-Wave Slot Antenna[2] have received considerable attention. Basically, the two structures mentioned consist of conventional cylindrical waveguides in which the internal fields are coupled to the exterior, or free space region, by means of a continuous axial slot. In the homogeneous case, the propagation constant is complex and the traveling-wave, which has a phase velocity greater than that of free space, is attenuated with distance along the guide. A useful means of radiation pattern control is available since the aperture illumination may be suitably adjusted by varying the transverse geometry of the waveguide and the width of the axial slot as functions of distance along the guide.

The transversely slotted waveguide[3] of Fig. 1 differs electrically from the traveling-wave slot antenna only in the type of coupling mechanism which is employed. The resulting differences between the characteristics of the two structures are in the modes of propagation and in the type of polarization of the radiated fields. In addition, the treatment of the transversely slotted, thin-walled waveguide given here may be viewed as a complement to the theoretical work by Elliott on the serrated rectangular waveguide,[4] where infinitesimally spaced transverse slots are cut into, and extend completely across, a broad wall of variable thickness.

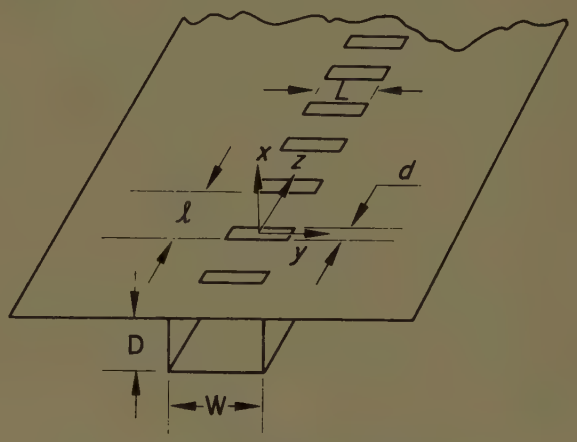


Fig. 1—Transversely slotted waveguide geometry.

Discussion

In order to simplify the analysis of the problem, the following idealizations have been made:

1) The waveguide is infinitely long and is immersed in an infinite ground plane.
2) The slot length-to-width ratio, L/d (see Fig. 1), is sufficiently great and the d/λ sufficiently small that only the z component of tangential electric field is appreciable.
3) The thickness of the slotted wall is infinitesimal.

In addition, μ and ϵ are taken equal to their free space values throughout, although a consideration of the case where the internal and external regions are individually homogeneous requires no essential extension of the theory.

As a consequence of the periodicity of the structure, the Floquet theorem[5] applies, and all fields are of the form[6]

$$F(x, y, z) = G(x, y, z)e^{-i\gamma z} \tag{1}$$

where γ is independent of the coordinates and G satisfies the relation

$$G(x, y, z) = G(x, y, z - l). \tag{2}$$

* Manuscript received by the PGAP, November 17, 1958; revised manuscript received, April 20, 1959. This work was supported by the Wright Air Dev. Command under Contract AF 33(616)-3220 and was submitted to the University of Illinois, Urbana, Ill., as partial fulfillment of the requirements for the Ph.D. degree.

† Ground Systems Group, Hughes Aircraft Co., Fullerton, Calif.; formerly at Antenna Lab., Univ. of Illinois, Urbana, Ill.

[1] W. Rotman, "The Channel Guide Antenna," Rep. No. E5054, U. S. Air Force Cambridge Res. Labs., Cambridge, Mass.; January, 1950.

[2] V. H. Rumsey, "Traveling-wave slot antennas," *J. Appl. Phys.*, vol. 24, p. 1366; November, 1953.

[3] R. F. Hyneman, "Closely Spaced Transverse Slots in Rectangular Waveguide," TR No. 14, Antenna Lab., University of Illinois. Urbana, Ill.; December, 1956.

[4] R. S. Elliott, "Serrated waveguide—Part I: theory," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-5, p. 270; July, 1957.

[5] L. Brillouin, "Wave Propagation in Periodic Structures," Dover Publications, Inc., New York, N. Y., p. 139; 1953.

[6] The notation $e^{-i\gamma z}$, has been employed in place of the more conventional $e^{-\gamma z}$ because of a slight simplification in subsequent equations.

All field components may be written in terms of E_t, the tangential electric aperture field. The tangential field in the cell of a typical slot, centered at the origin, is given by

$$E_t = zE_z(0, y, z) = zg(y, z)\begin{Bmatrix} 0 \le |y| \le \frac{L}{2} \\ 0 \le |z| \le \frac{d}{2} \end{Bmatrix}$$
$$= 0 \qquad \begin{Bmatrix} \frac{L}{2} \le |y| \\ \frac{d}{2} \le |z| \le \frac{l}{2} \end{Bmatrix} \tag{3}$$

where $g(y, z)$ is an unknown scalar function of position and z is the unit vector in the z direction. Since $E_z e^{+j\gamma z}$ is periodic,

$$E_z(0, y, z) = \sum_{n=-\infty}^{\infty} a_n(y) e^{-j\gamma_n z} \tag{4}$$

where

$$\gamma_n = \gamma - \frac{2n\pi}{l},$$
$$a_n(y) = \frac{1}{l}\int_{-d/2}^{d/2} g(y, z) e^{j\gamma_n z} dz, \qquad 0 \le |y| \le \frac{L}{2}$$
$$= 0 \qquad \frac{L}{2} \le |y|.$$

Because E_t has no y component, and in view of the symmetrical boundary considerations, all fields are TE with respect to the y axis. With f, the magnetic scalar potential, defined by

$$E = -\nabla \times yf = y \times \nabla f, \tag{5}$$

the scalar wave equation yields (with use of the $e^{+j\omega t}$ time convention),

$$H_y = \frac{1}{j\omega\mu}\left(k^2 + \frac{\partial^2}{\partial y^2}\right) f,$$
$$H_z = \frac{1}{j\omega\mu}\frac{\partial^2 f}{\partial y \partial z} \tag{6}$$

where $k^2 = \omega^2\mu\epsilon$. Thus, continuity of tangential H at each slot is enforced by the condition

$$f_{\text{int}}(0, y, z) = f_{\text{ext}}(0, y, z) \tag{7}$$

where f_{int} and f_{ext} are the scalar potentials, respectively, in the internal and external regions. With the use of solutions of the scalar wave equation which satisfy the boundary condition of zero tangential E at the conducting surfaces, and the radiation condition at infinity, (4), (5), and (7) give

$$\int_{-L/2}^{L/2}\int_{-d/2}^{d/2} g(y, z) K(y, z \mid y', z') dy' dz' = 0 \tag{8}$$

where

$$K(y, z \mid y', z') = \sum_{n=-\infty}^{\infty}\left\{\frac{2}{lW}\sum_{m=1}^{\infty}{}' \frac{\cot K_{m,n} D}{K_{m,n}} \cos\frac{m\pi y}{W}\cos\frac{m\pi y'}{W} e^{-j\gamma_n(z-z')} + j\frac{1}{2l} H_0^{(2)}(\beta_n |y - y'|) e^{-j\gamma_n(z-z')}\right\},$$

with

$$K_{m,n} = \sqrt{k^2 - \left(\frac{m\pi y}{W}\right)^2 - \gamma_n^2},$$
$$\beta_n = \sqrt{k^2 - \gamma_n^2}$$

and where the prime on sigma ($\sum'$) denotes summation over the odd numbered terms only.

Expression (8) is an integral equation for the determination of the aperture field and γ. Since the equation is not of standard form and no technique for exact solution exists, a variational solution for an approximate value of γ must be sought. Unfortunately, the non-symmetrical (in z) property of the kernel in (8) appears to make impossible the establishment of a variational principle for the general case where γ is complex. However, approximate formulas for γ may be obtained which do have stationary properties for either of the restricted conditions that γ be real, or that the slot spacing be infinitesimal.

For example, assume γ real and $\gamma > k$. Then it can be shown[3,7] that K is Hermitian, *i.e.*,

$$K(y, z \mid y', z') = K^*(y', z' \mid y, z),$$

and that the expression

$$\int_{-d/2}^{d/2}\int_{-d/2}^{d/2}\int_{-L/2}^{L/2}\int_{-L/2}^{L/2} g(y, z) g^*(y', z') \cdot K(y, z \mid y', z') dy dy' dz dz' = 0 \tag{9}$$

gives a value of γ which is stationary for small variations of an assumed g about a correct solution of (8).

For the uniform, longitudinally slotted waveguide case, Rumsey[2] has shown that the expression

$$\int_{-L/2}^{L/2} (E_{a_x} J_{a_z} - E_{a_y} J_{a_y}) dy = 0, \tag{10}$$

(where L is the width of the longitudinal slot, E_a is the transverse distribution of the assumed aperture field, and where J_a is calculated from the resulting discontinuity of magnetic field in the aperture which fits

[7] P. M. Morse and H. Feshbach, "Methods of Theoretical Physics," Section 9.4, McGraw-Hill Book Co., Inc., New York, N. Y.; 1953.

E_a) is a second variational formula which has stationary properties for arbitrary complex γ. The left-hand side of (10) is the formula for the self reaction,[8] $\langle a, a\rangle$, of the traveling-wave source distribution, $J_a(y)e^{-i\gamma z}$, which when established in each slot supports the initially assumed slot field, E_a. Eq. (10) does not apply for the transversely slotted guide, since in this case the slot field is made up of an infinite number of traveling waves of differing propagation constants, γ_n.

A third aspect of the approximate method is brought out by considering an aperture of finite extent, S. In this case, the conventional form of the reciprocity theorem applies and the self reaction of the aperture source, J_a, is given by

$$\langle a, a\rangle = \int\int_S J_a \cdot E_a dS. \tag{11}$$

Eq. (11) also does not strictly apply to the transversely slotted waveguide, because the infinite extent of the aperture violates the condition of the reciprocity theorem that the source be of finite extent. Thus, although the relation

$$\int\int_{S'} E_{a_z} J_{a_z} dS' = \int\int_{S'} E_{a_z}[H_y^{\text{ext}} - H_y^{\text{int}}]dS' = 0, \tag{12}$$

obtained by applying (11) to the *infinite* aperture surface S', gives an approximate expression for γ in terms of the assumed E_{a_z}, the expression is not necessarily stationary. However, it can be shown that in the limit as $l/\lambda \to 0$, (12) reduces to (10), since in this case the aperture distribution becomes essentially continuous and the then trivial dependence on z may be eliminated. Thus γ derived from (12) is approximately stationary for small l/λ and complex γ, while that derived from (9) is stationary for real γ and arbitrary l/λ. There is some question as to whether (9) or (12) constitutes the better approximation for γ in the particular case under consideration. On the basis of physical reasoning, it would appear that (12) more nearly fits the actual circumstances since, in general, γ is complex and small values of l/λ are of major interest. In particular, it will be seen that the limiting case of infinitesimal l/λ is one of special interest.

A. Approximate Expression for γ

It is apparent that when the integration of (12) over any typical slot is set equal to zero, the condition that the integral over the infinite aperture be zero is also satisfied. In (3), the approximation to the slot field distribution is taken to be

$$g(y, z) = \cos \pi y/L, \qquad 0 \leq |y| \leq L/2. \tag{13}$$

Substitution into (12) with the use of (5) and (6), and integration over the slot centered at the origin gives

$$\frac{\cot \kappa D}{\kappa D} = \frac{\left(1 - \frac{L^2}{W^2}\right)^2 \quad \left(\frac{\gamma d}{2}\right)^2 \quad 1}{\cos^2 \frac{\pi L}{2W} \quad \sin^2 \frac{\gamma d}{2} \quad \left(1 - \frac{\lambda^2}{4W^2}\right)}$$

$$\cdot \sum_{n=-\infty}^{\infty} \frac{\sin^2\left(\frac{n\pi d}{l} - \frac{\gamma d}{2}\right)}{\left(\frac{n\pi d}{l} - \frac{\gamma d}{2}\right)^2}$$

$$\cdot \left\{ \sum_{m=1}^{\infty} \xi_{m,n} \left(\frac{m^2\lambda^2}{4W^2} - 1\right) \frac{\cos^2 \frac{m\pi L}{2W}}{\left(1 - \frac{m^2L^2}{W^2}\right)^2} \frac{\cot \kappa_{m,n} D}{\kappa_{m,n} D} \right.$$

$$+ -j \frac{\pi^2}{16} \frac{W}{k^2 L^2 D} \int_{-L/2}^{L/2} \int_{-L/2}^{L/2} \cos \frac{\pi y}{L} \cos \frac{\pi y'}{L} \left(k^2 + \frac{\partial^2}{\partial y^2}\right)$$

$$\left. \cdot H_0^{(2)}(\beta_n | y - y' |) dy dy' \right\}, \tag{14}$$

where

$$\xi_{m,n} \begin{cases} = 0 & \text{for } m = 1 \quad \text{when } n = 0 \\ = 1 & \text{otherwise} \end{cases}$$

$$\kappa \equiv \kappa_{1,0}.$$

The set of double integrals making up the final term of (14) may be written[2] in terms of the form:

$$g_n = 4\pi \left(\frac{L}{\lambda}\right)^2 \int_0^{\pi} H_0^{(2)}\left(\frac{\beta_n L r}{\pi}\right)$$

$$\cdot \left[\left(1 - \frac{\lambda^2}{4L^2}\right) \cos r + \frac{1}{\pi}\left(1 + \frac{\lambda^2}{4L^2}\right) \sin r\right] dr. \tag{15}$$

For $n = 0$ the power series representation for $H_0^{(2)}$ may be substituted and the resulting series evaluated term by term. For $n \neq 0$ and for l/λ sufficiently small that

$$|\beta_n L| > 3\pi$$

(15) may be approximately evaluated in closed form (Appendix).

As a further aid in the solution of (14), the left-hand side may be easily written in terms of a perturbation of γ. Let

$$\gamma = \gamma(0) + \delta(\gamma) = \sqrt{k^2 - \frac{\pi^2}{W^2}} + \delta(\gamma); \tag{16}$$

then

$$\kappa = \sqrt{k^2 - \frac{\pi^2}{W^2} - \gamma^2} \simeq \sqrt{2\gamma(0)\delta(\gamma)}. \tag{17}$$

By use of the first two terms of the power series expansion for $\cot z/z$, the expression

$$\frac{\cot \kappa D}{\kappa D} \simeq \frac{1}{2\gamma(0) D^2 \delta(\gamma)} \tag{18}$$

[8] V. H. Rumsey, "Reaction concept in electromagnetic theory," *Phys. Rev.*, vol. 94, p. 1483; June 15, 1954.

is obtained. For cases of small perturbation, (18) used in conjunction with an iterative technique provides a straightforward method of solution of (14).

Unfortunately, the solution of (14) is still quite tedious, because of the relatively slow convergence of the doubly infinite series. The formula is considerably simplified for either of the two special limiting cases, a) for $L=W$, where the slots extend completely across the width of the waveguide, and b) for infinitesimal slot spacing, *i.e.*, as $l/\lambda \rightarrow 0$. It turns out that the results for these special cases are sufficiently general to be of considerable practical interest.

B. Theoretical and Experimental Results for the Case of $L=W$

For $L=W$, only the $m=1$ terms of (14) are non-zero. Since in this case the perturbation of γ caused by the presence of the slots generally is relatively great, the iterative technique of solution is of little value and a direct solution of (14) is necessary. Because of the transcendental character of the equation, an automatic digital computer was employed in the solutions leading to the results in Figs. 2 to 4.

Fig. 2 gives typical variations with frequency of the attenuation constant, α, for various slot widths and spacings. Of some interest is the behavior of α in the vicinity of the unperturbed cutoff frequency since the result shown contrasts with that obtained in the axially slotted waveguide case. In the latter case, α increases indefinitely as the frequency decreases towards cutoff. The effect in the transversely slotted case may be explained on the following basis. Exactly at the cutoff frequency, the fields in the unperturbed (closed) waveguide would be TEM with respect to the y axis. Since in the transversely slotted case the long dimensions of the slots are then parallel with the direction of current flow on the broad face of the guide at this frequency, it is to be expected that the perturbation of the field components would be small and the radiation losses minimal.

At frequencies below unperturbed cutoff, the fields decay more rapidly than in the unperturbed waveguide as a consequence of the radiation losses. In addition, the waveguide velocity ratio, c/v, where c is the free space velocity, is finite in this region instead of remaining zero as in the lossless, unperturbed case. However, the signs of c/v and α are opposite, so that the direction of decreasing phase is opposite to the direction of decreasing amplitude. A comparison of the perturbed and unperturbed c/v ratios is given in Fig. 3.

Of further interest in Fig. 2 is the relative insensitivity of α to changes in l/d. The broken curve, which corresponds to an impractically small value of d, represents only a slight decrease in the average value of α. A moderate degree of control over α is available, however, through the variation of l/λ. In the typical results given in Fig. 4, the maximum range of α for variation of l/λ alone is seen to be approximately 10:1. In addition, since α increases without limit as D is decreased, the range of α can be correspondingly increased by

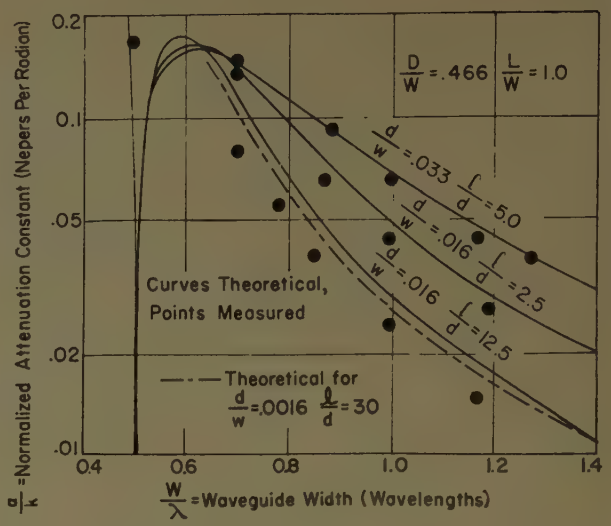


Fig. 2.—Typical variation of attenuation constant for $L=W$.

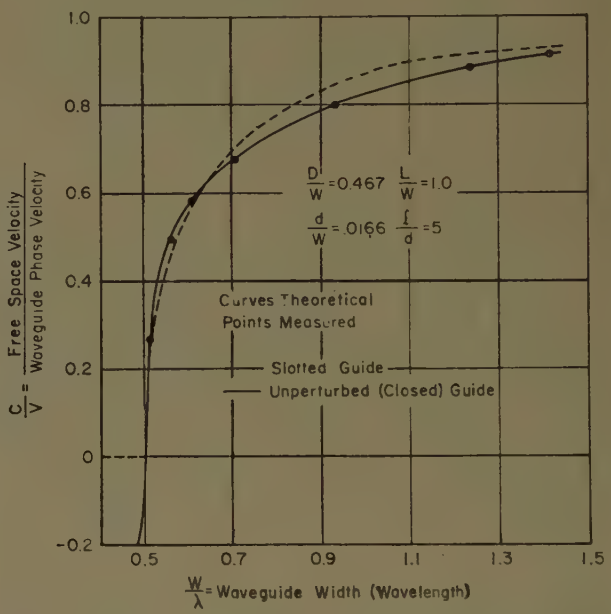


Fig. 3.—Comparison of perturbed and unperturbed velocity ratios for guide of Fig. 2.

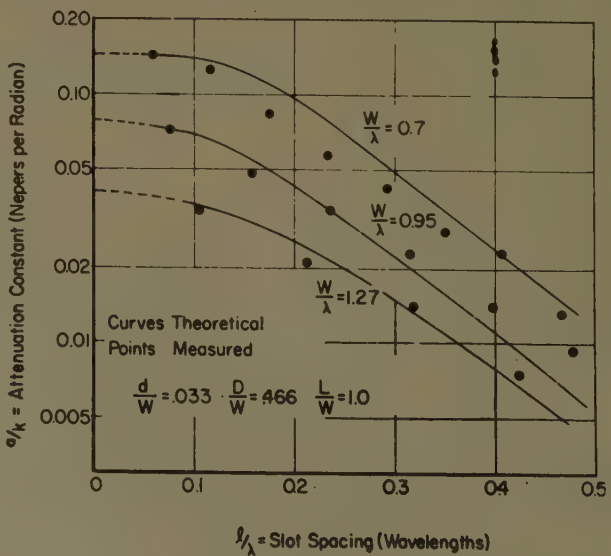


Fig. 4.—Control of attenuation constant by slot spacing with $L=W$.

allowing a simultaneous variation of D. The variation c/v as a function of the slot spacing is given in Fig. 5.

C. Theoretical and Experimental Results for Infinitesimal l/λ and Variable L/W

For infinitesimal slot spacing, only the $n=0$ terms of (14) are non-zero. For cases of small slot length, the perturbation of γ from the closed waveguide is relatively small and the iterative technique of solution may be profitably employed. Typical results for a waveguide of standard aspect ratio are given in Figs. 6 and 7. Of some interest is the fact that α is practically insensitive to variations of L above about $\lambda/2$. Effective control over α is obtained only with relatively small values of L/λ.

The results of this idealized analysis based on an assumed infinitesimal slot spacing appear to give good agreement with experimental results obtained with actual slot spacings of the order of one-tenth wavelength or less and with $l/d \leq 2$.

D. Surface Wave Solution

It can be shown[3] that under certain conditions, (14) admits to a surface wave solution (*i.e.*, where γ is real and greater than k) in addition to the solution corresponding to the perturbed, fundamental mode. Surface wave propagation is possible over a band of frequencies with upper and lower cutoffs determined by the guide and slot parameters. For infinitesimal slot spacing, the lower cutoff frequency is always zero. In addition, for the particular case of $L=W$, the upper cutoff for the lowest order mode, $E_z \simeq \cos \pi y/L$, is (to the limits of accuracy imposed by the approximate expressions) at $L=\lambda/2$.

A typical plot of c/v vs W/λ for $L/W=1$ obtained from the computer solution of (14) is given in Fig. 8. No results are given for the case of $L/W<1$ since for the surface wave the perturbation of γ from that of the closed waveguide case is relatively great and the iterative technique is thus of little value, while the direct

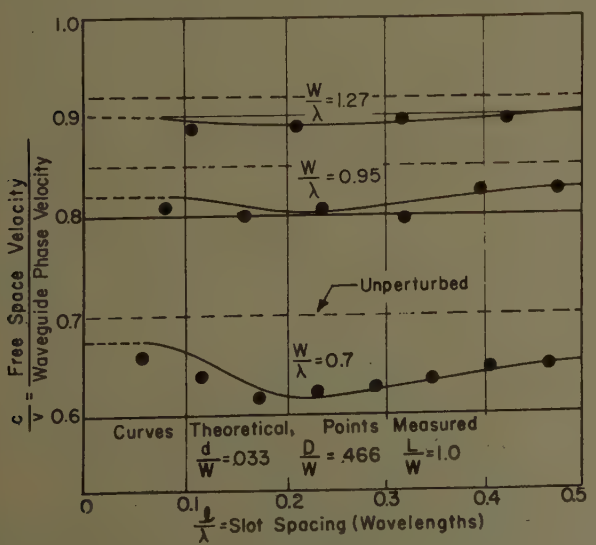


Fig. 5.—Perturbation of velocity ratio for same cases as Fig. 4.

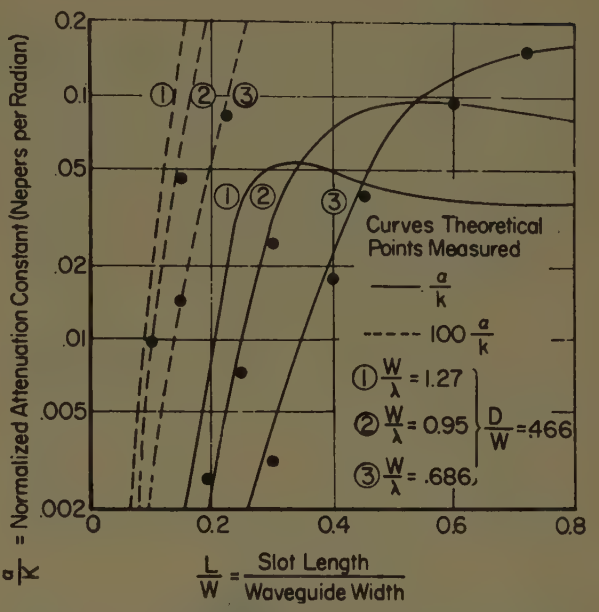


Fig. 6.—Effect of slot length on attenuation constant for infinitesimally spaced slots.

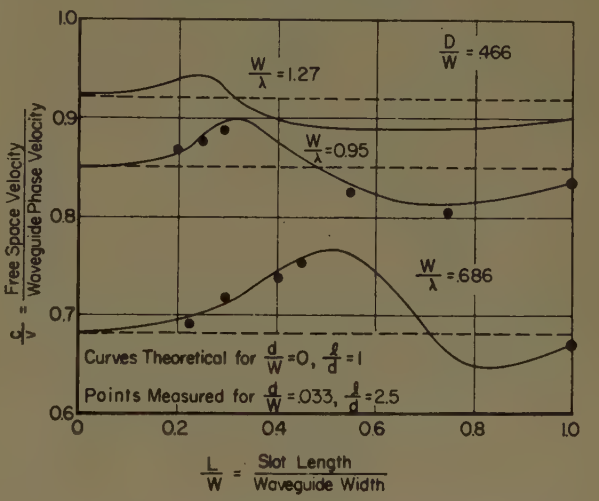


Fig. 7.—Variations in velocity ratios for same cases as Fig. 6.

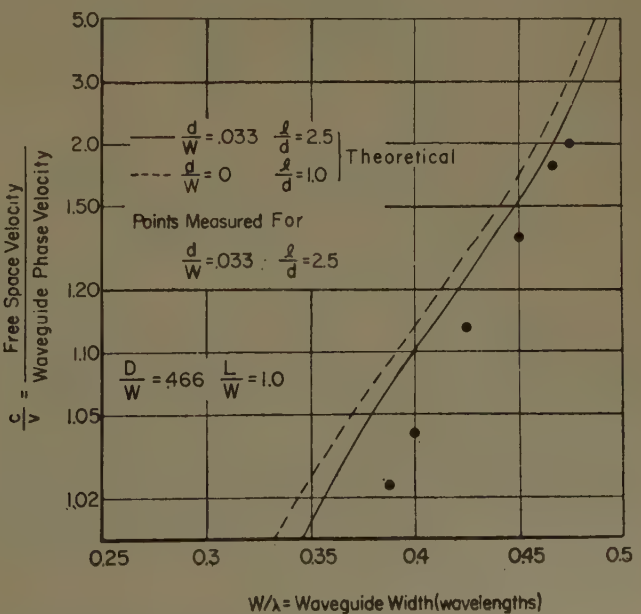


Fig. 8.—Typical variation in velocity ratio for surface wave mode of propagation.

solution of (14) is quite tedious. However, for infinitesimal slot spacing an asymptotic expression valid as $\omega \to 0$, $c/v \to 1$ is easily developed.

For $c/v \simeq 1$, only the first term of the power series expansion for $H_0^{(2)}$ is necessary in (15). Performing the integration gives (see Appendix):

$$g_0 \simeq -j\frac{2}{\pi}\left\{1.27\left[c + \ln\frac{\sigma}{k}\frac{L}{\lambda}\right] - 0.807 + 1.215\frac{\lambda^2}{4L^2}\right\} \quad (19)$$

where

$$\sigma = +j\beta = k\sqrt{\left(\frac{c}{v}\right)^2 - 1} \simeq k\sqrt{2\left(\frac{c}{v} - 1\right)},$$

c = Euler's constant = 0.577 · · · ·

Further, as

$$\frac{c}{v} \to 1, \qquad \kappa \to \sqrt{-\left(\frac{\pi D}{W}\right)^2}$$

and

$$\frac{\cot \kappa D}{\kappa D} \to -\frac{\coth \frac{\pi D}{W}}{\frac{\pi D}{W}}.$$

Substitution into (14) with $l/\lambda \to 0$, $l/d = 1$, and a subsequent rearrangement of terms gives

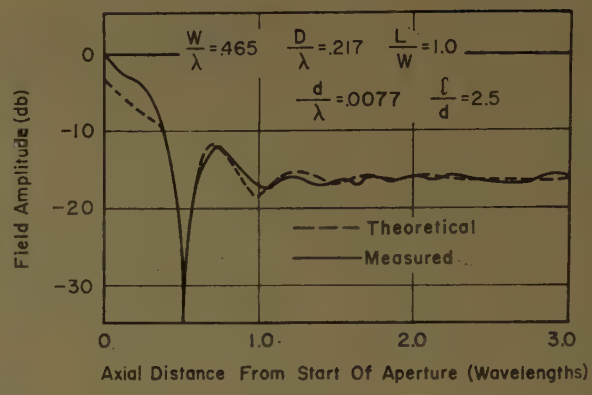
$$\frac{c}{v} \simeq 1 + \exp\left\{0.808 - \frac{2\lambda S}{W} - 1.91\frac{\lambda^2}{4L^2} + n\frac{\lambda^2}{4L^2} + 4\left(1 - \frac{\lambda^2}{4W^2}\right)\frac{\cos^2\frac{\pi L}{2W}}{\left(1 - \frac{L^2}{W^2}\right)^2}\coth\frac{\pi D}{W}\right\} \quad (20)$$

where

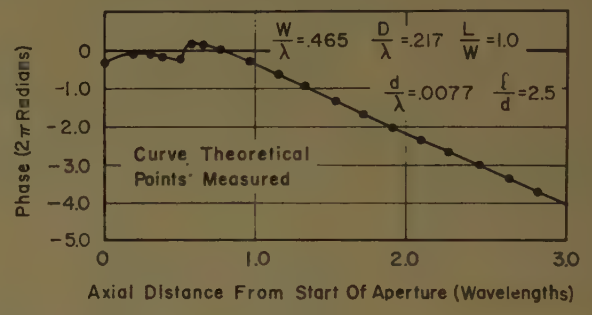
$$S = \sum_{m=3}^{\infty}{}' \frac{\cos^2\frac{m\pi L}{2W}}{\left(1 - \frac{m^2L^2}{W^2}\right)^2}\left(\frac{m^2\lambda^2}{4W^2} - 1\right)\frac{2W}{m\lambda}.$$

In general, both the fundamental and surface wave modes are excited by an incident $TE_{1,0}$ field when $L < \lambda/2$. For a typical condition of $L = W < \lambda/2$, the measured variations of phase and amplitude with axial distance along the waveguide are given in Fig. 9. The computed curves in this figure were obtained by taking a solution of (14) for γ for the rapidly damped fundamental mode,

$$\frac{\gamma}{k} = \frac{c}{v} - j\frac{\alpha}{k} = -0.190 - j0.438,$$



(a)



(b)

Fig. 9.—Axial phase-amplitude distributions with fundamental and surface wave modes present, $W < \lambda/2$. (a) amplitude; (b) phase.

and the measured value of

$$\frac{\gamma}{k} = \frac{c}{v} = 1.77$$

for the surface wave. Relative phase and amplitude of these two fields were then adjusted to give the best fit with the measured results. The poor agreement at the beginning of the aperture is possibly due to the presence of higher order evanescent modes.

For cases where $L < \lambda/2 < W$, excitation of the surface wave mode relative to that of the fundamental mode may be appreciable, reaching a maximum in the range of $0.25\lambda < L < 0.4\lambda$. For those antenna applications where pure fundamental mode propagation is required, the existence of this surface wave mode thus restricts L to values less than the order of 0.2λ for waveguide of standard aspect ratio. While a decrease in D results in an increase in α for a given L/λ, the relative excitation of the surface wave also appears to be thereby increased, so that an effective upper limit on the value of α, obtainable by varying L/λ, is still imposed. Where higher rates of attenuation than are thus available are required for the desired aperture illumination function, values of $L > \lambda/2$ should be used, *e.g.*, $L = W$, with α then being controlled by variation in the slot spacing. There is a small degree of overlap in the ranges of α available by the two means of control.

Experimental Measurements

The presence of the surface wave also causes some difficulty in the verification of the predicted attenuation

constant. In cases where the excitation of this mode was appreciable, the attenuation constant was obtained by probing the field as a function of distance along the guide. The slope of the straight line portion of a db plot of the field strength was used to give an accurate value of α. In cases where the attenuation constant was small, and the surface wave excitation negligible, the attenuation was determined from VSWR measurements at the input of the slotted guide. In order to account for the waveguide wall losses and the reflection produced at the beginning of the slots, measurements were made of various lengths of slotted sections for each case reported. In all cases, the phase velocity was determined from field measurements along the guide.

There is some disrepancy between curve 3 of Fig. 6 and the experimental results given by Kelly and Elliott[9] in their Figs. 5 and 6. It is believed that this is due to the appreciable wall thickness employed in their measurements. For the experimental results in the present paper, the slots were milled completely across the top wall of the guide and the transverse slot length was then adjusted by covering over part of the slotted broad wall with adhesive-backed metal foil, thus giving an effective wall thickness roughly that of the foil itself.

Conclusion

It has been shown that the propagation constant of the transversely slotted waveguide can be effectively controlled through variation of the guide and slot parameters with considerable independence of control over the attenuation constant and phase velocity available. Experimental results show good agreement with the theory.

A further investigation of the surface wave supporting properties of the structure is indicated. Of importance are methods of efficient surface wave excitation which do not depend on external launching devices such as waveguide horns. Alternatively, the characteristics of different periodic coupling mechanisms, *e.g.*, closely spaced, circular holes, should be studied from the standpoint of attempting to decrease the surface wave excitation when the perturbed fundamental mode of propagation is desired.

Finally, the effects of finite wall thicknesses on γ should also be considered. This parameter is of particular importance in cases of small slot length, where the slot region becomes a section of below cutoff waveguide.

Appendix—Evaluation of Eq. (15)

For $n \neq 0$, $l < \lambda/2$, then the parameter σ_n, where

$$\sigma_n = +j\beta_n = k\sqrt{\left(\frac{n\lambda}{l} - \frac{\gamma}{k}\right)^2 - 1},$$

is approximately real and positive. Since

$$H_0^{(2)}(\beta_n | y - y' |) = j\frac{2}{\pi} K_0(\sigma_n | y - y' |),$$

then with $a = \pi/\sigma_n L$, (15) becomes

$$\begin{aligned} g_n &= j\frac{2a}{\pi}\int_0^{\pi/a} K_0(r)\left[\left(1 - \frac{\lambda^2}{4L^2}\right)\left(1 - \frac{ar}{\pi}\right)\cos ar \right. \\ &\qquad \left. + \frac{1}{\pi}\left(1 + \frac{\lambda^2}{4L^2}\right)\sin ar\right] dr \\ &= j\frac{2a}{\pi}\left\{\int_0^{\infty}[\ ]dr - \int_{\pi/a}^{\infty}[\ ]dr\right\} \\ &= j\frac{2a}{\pi}\{g_{n_1} - g_{n_2}\}. \end{aligned}$$

Formulas for the evaluation of g_{n_1} are available:[10]

$$\int_0^{\infty} K_0(r)\cos ardr = \frac{\pi}{2}\frac{1}{\sqrt{1+a^2}}$$

$$\int_0^{\infty} K_0(r)\sin ardr = \frac{\operatorname{arc\,sinh} a}{\sqrt{1+a^2}}$$

$$\int_0^{\infty} rK_0(r)\cos ardr = \frac{1}{2}\sqrt{\frac{\pi}{2}}[\Gamma(2)]^2$$

$$\cdot\left\{\frac{P_{-1/2}^{-3/2}(ja)}{\sinh^{3/2}[\operatorname{arc\,cosh} ja]} + \frac{P_{-1/2}^{-3/2}(-ja)}{\sinh^{3/2}[\operatorname{arc\,cosh}(-ja)]}\right\}$$

where P is the generalized Legendre polynomial. The assumption that

$$|a| < \tfrac{1}{2}$$

enables use of the approximations

$$\sqrt{1+a^2} \simeq 1 + \frac{a^2}{2},$$

$$\operatorname{arc\,cosh}(\pm ja) \simeq a - \frac{a^2}{2} \pm j\frac{\pi}{2},$$

$$\sinh^{3/2}[\operatorname{arc\,cosh} \pm ja] \simeq e^{\pm j(3\pi/4)}\left(1 + \frac{3}{4}a^2\right)$$

$$\begin{aligned} &P_{-1/2}^{-3/2}(\pm ja) \\ &= \frac{1}{\Gamma\left(\frac{5}{2}\right)}\left(\frac{\pm ja - 1}{\pm ja + 1}\right)^{3/4} F\left(\frac{1}{2}, \frac{1}{2}; \frac{5}{2}; \frac{1 \mp ja}{2}\right) \\ &\simeq \frac{1}{\Gamma\left(\frac{5}{2}\right)} e^{\pm j(3\pi/4)}[1 \mp j1.5a - 1.125a^2](1.085 \mp 0.165ja). \end{aligned}$$

[9] K. C. Kelly and R. S. Elliott, "Serrated waveguide—Part II: experiment," IRE Trans. on Antennas and Propagation, vol. AP-5, p. 276; July, 1957.

[10] G. N. Watson, "Theory of Bessel Functions," The Macmillan Co., New York, N. Y., p. 338; 1944.

Collecting significant terms gives

$$g_{n_1} \simeq \frac{\pi}{2}\left[\left(1-\frac{\lambda^2}{4L^2}\right)+\frac{\lambda^2}{4L^2}\,\frac{4}{\pi^2}\,a\right].$$

In a straightforward manner, it can be shown that

$$|g_{n_2}| \ll |g_{n_1}|,$$

and thus

$$g_n \simeq j\,\frac{2a}{\pi}\,g_{n_1}.$$

For $n=0$, substitution of the power series for $H_0^{(2)}(z)$ into (15) and subsequent integration gives

$$g_0 = \sum_{i=0}^{\infty}(-1)^i\left(\frac{\beta}{\kappa}\,\frac{L}{\lambda}\right)^{2i} \cdot\left\{\left[1-j\frac{2c}{\pi}-j\frac{2}{\pi}\ln\left(\frac{\beta}{\kappa}\,\frac{L}{\lambda}\right)\right]A_{2i}+jB_{2i}\right\}$$

where

$$A_{2i} = a_{2i}+\frac{\lambda^2}{4L^2}\,b_{2i},$$

$$B_{2i} = \frac{2}{\pi}\left[A_{2i}\sum_{r=1}^{i}\frac{1}{r}-c_{2i}-\frac{\lambda^2}{4L^2}\,d_{2i}\right],$$

$$c = \text{Euler's constant} = 0.5772\cdots,$$

and where

$$\begin{Bmatrix} a_{2i} \\ b_{2i} \end{Bmatrix} = \frac{1}{(i!)^2}\int_0^{\pi}\frac{r^{2i}\sin r\,dr}{\pi} \cdot\begin{Bmatrix} + \\ - \end{Bmatrix}\frac{1}{(i!)^2}\int_0^{\pi} r^{2i}\left(1-\frac{r}{\pi}\right)\cos r\,dr,$$

$$\begin{Bmatrix} c_{2i} \\ d_{2i} \end{Bmatrix} = \frac{1}{(i!)^2}\int_0^{\pi}\frac{r^{2i}}{\pi}\sin r\ln r\,dr \cdot\begin{Bmatrix} + \\ - \end{Bmatrix}\frac{1}{(i!)^2}\int_0^{\pi} r^{2i}\left(1-\frac{r}{\pi}\right)\cos r\ln r\,dr.$$

ACKNOWLEDGMENT

The author wishes to express his appreciation to Prof. V. H. Rumsey and Dr. R. H. Duhamel for their guidance and their many helpful suggestions. He is also indebted to Mrs. Allen Blankfield who performed most of the computations, and to J. J. Stafford who performed many of the experimental measurements.

Generalizations of Spherically Symmetric Lenses*

SAMUEL P. MORGAN†

Summary—**The purpose of this paper is to generalize the solutions of some spherically symmetric lens and lens-reflector problems recently treated by Kay. The original problem was to find a variable-index structure, with a point source at its surface or at infinity, which would produce a beam of finite angular width, having a prescribed variation of intensity with angle. It is shown that a prescribed exit beam can be obtained from a point source at any given distance from the lens, and that the index of refraction may be specified more or less arbitrarily in the outer part of the lens. A special case is solved in terms of tabulated functions.**

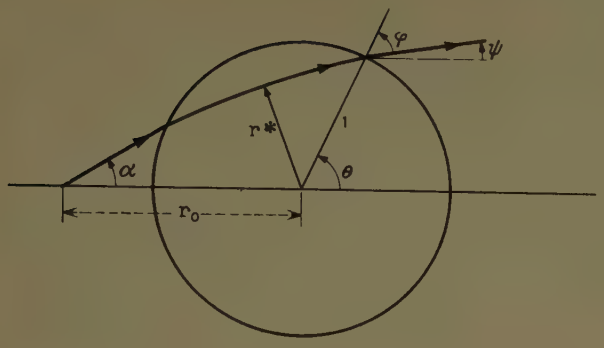
INTRODUCTION

IN AN interesting paper, Kay[1] has discussed various generalizations of the cylindrical or spherical Luneberg lens. The lenses treated by Kay have a point source at the surface or at infinity, and form a beam of finite angular width, having a prescribed variation of intensity with angle. In each case he finds a single expression for the index of refraction of the lens as a function of the radial coordinate. We shall show that Kay's problems can easily be solved in the more general case where the incident beam comes from a point source anywhere outside the lens, and that the index of refraction may be chosen arbitrarily, subject to simple restrictions, in an outer shell of any desired thickness less than the radius of the lens. The method of solution is very similar to that given in a previous paper[2] on the original Luneberg lens problem, and the exposition will be correspondingly brief.

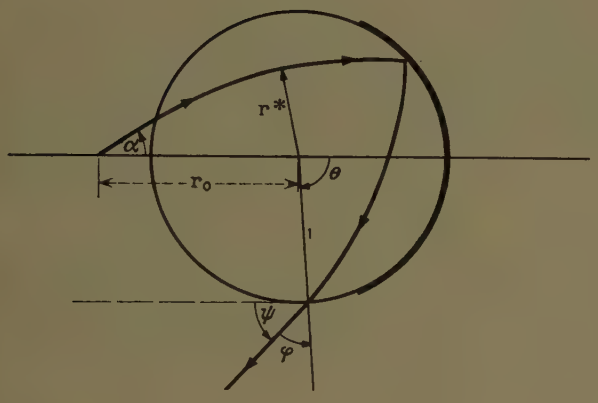
Consider the two spherical structures shown in cross section in Fig. 1. In each case the lens radius is taken as the unit of length. Fig. 1(a) represents a broad-beam

* Manuscript received by the PGAP, April 6, 1959.
† Bell Telephone Labs., Inc., Murray Hill, N. J.
[1] A. F. Kay, "Spherically symmetric lenses," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-7, pp. 32–38; January, 1959.

[2] S. P. Morgan, "General solution of the Luneberg lens problem," *J. Appl. Phys.*, vol. 29, pp. 1358–1368; September, 1958.



(a)



(b)

Fig. 1—(a) Broad-beam Luneberg lens. (b) Broad-beam Luneberg reflector.

Luneberg lens, in which rays from a point source at a distance $r_0 \geq 1$ from the center of the lens are formed into a rotationally symmetric exit beam with a prescribed intensity in each angular direction ψ. Fig. 1(b) shows a broad-beam Luneberg reflector, in which a similar exit beam is formed by rays which have passed through the lens twice, with an intervening reflection at the back surface. Obviously, if the Luneberg reflector is to be used with an arbitrary source direction, then its surface must be only partially silvered. In that case we shall neglect the variation of reflection and transmission coefficients with angle.

We shall not discuss the broad-beam Eaton-Lippmann lens, which is the third type treated by Kay. In an Eaton-Lippmann lens the rays are not reflected but are refracted through an angle so great that they emerge in a direction opposite to the direction of incidence. The scalar theory of the Eaton-Lippmann lens is quite analogous to the theory of the lenses shown in Fig. 1; but as Kay points out, for a spherical Eaton-Lippmann lens the polarization relationships are such that scalar theory is completely inadequate to describe the back-scattering cross section.

Formal Solution

In Fig. 1, let α be the angle which a typical incident ray makes with the axis, ψ the angle from the axis to the direction of the exit ray, ϕ the angle from the exit ray to the radius vector at the exit point, and θ the polar angle corresponding to the exit point. Note that θ is a positive angle in Fig. 1(a), and a negative angle in Fig. 1(b). Each ray corresponds to a different value of the parameter κ defined by

$$\kappa = r_0 \sin \alpha, \qquad 0 \leq \kappa \leq 1. \tag{1}$$

It is clear that the angles α, ψ, ϕ, and θ may all be regarded as functions of κ.

Using the well-known equation[2] for the path of a light ray in a radially symmetric refracting medium, it is easy to derive the integral equation which must be satisfied by the index of refraction of the lens. The result for the broad-beam Luneberg lens of Fig. 1(a) is

$$\int_{r^*(\kappa)}^{1} \frac{\kappa dr}{r[\rho^2(r) - \kappa^2]^{1/2}} = f(\kappa), \qquad 0 \leq \kappa \leq 1, \tag{2}$$

and for the broad-beam Luneberg reflector of Fig. 1(b),

$$\int_{r^*(\kappa)}^{1} \frac{\kappa dr}{r[\rho^2(r) - \kappa^2]^{1/2}} = \frac{1}{2} f(\kappa), \qquad 0 \leq \kappa \leq 1, \tag{3}$$

where in both cases,

$$f(\kappa) = \frac{1}{2}\left[\pi + \sin^{-1}\frac{\kappa}{r_0} - \sin^{-1}\kappa - \theta(\kappa)\right]. \tag{4}$$

The function $\rho(r)$ is defined in terms of the refractive index $n(r)$ by

$$\rho(r) = rn(r); \tag{5}$$

and the "turning radius" $r^*(\kappa)$ is a root of

$$\rho(r) = \kappa. \tag{6}$$

The polar exit angle $\theta(\kappa)$ is assumed to be a continuously increasing function of κ. It will be determined in the next section in terms of the distributions of intensity in the incident and emergent beams.

By the same argument as for an ordinary Luneberg lens,[2] it follows that $\rho(r)$ may be taken as an arbitrary function greater than unity in the outer part of the lens. Thus we may let

$$\rho(r) = P(r) \geq 1, \qquad n(r) = P(r)/r \geq 1/r, \tag{7}$$

for a $\leq r \leq 1$, where $a > 0$ is the least upper bound of the values of r for which $\rho(r) < 1$. In the range $0 \leq r \leq a$, the function $\rho(r)$ is determined by the integral equation (2) or (3). If a solution exists, it must be a continuous, strictly increasing function of r, with $\rho(0) = 0$ and $\rho(a - 0) = 1$.

Consider now the broad-beam Luneberg lens shown in Fig. 1(a). If the marginal ray ($\kappa = 1$) is to make an angle ψ_0 with the positive x-axis after refraction through the lens, the refractive index of the outer shell must satisfy the condition

$$\int_a^1 \frac{dr}{r[P^2(r) - 1]^{1/2}} \leq \frac{1}{2}\left[\sin^{-1}\frac{1}{r_0} - \psi_0\right]. \tag{8}$$

Since the integral on the left cannot be negative, we have the obvious requirement that

$$\psi_0 \leq \sin^{-1}\frac{1}{r_0}. \tag{9}$$

The solution of (2) by a now-familiar technique[2] leads to the determination of the index of refraction in the inner region, $0 \leq r \leq a$, via the parametric equations

$$n = (1/a)\exp\left[\omega(\rho, r_0) + \omega(\rho, 1) - \Omega(\rho) - \Theta(\rho)\right],$$
$$r = \rho/n, \tag{10}$$

where the parameter ρ runs from 0 to 1. The definitions of the functions appearing in (10) are as follows:

$$\omega(\rho, s) = \frac{1}{\pi}\int_\rho^1 \frac{\sin^{-1}(\kappa/s)}{(\kappa^2 - \rho^2)^{1/2}}\,d\kappa, \quad 0 \leq \rho \leq 1, \quad s \geq 1; \tag{11}$$

$$\Omega(\rho) = \frac{2}{\pi}\int_\rho^1 \left[\int_a^1 \frac{dr}{r[P^2(r) - 1]^{1/2}}\right]\frac{\kappa\,d\kappa}{(\kappa^2 - \rho^2)^{1/2}}, \quad 0 \leq \rho \leq 1; \tag{12}$$

$$\Theta(\rho) = \frac{1}{\pi}\int_\rho^1 \frac{\theta(\kappa)\,d\kappa}{(\kappa^2 - \rho^2)^{1/2}}, \qquad 0 \leq \rho \leq 1. \tag{13}$$

The function $\omega(\rho, s)$ was introduced by Luneberg,[3] who did not, however, tabulate any numerical values. Fletcher, Murphy, and Young[4] developed a series for $\omega(\rho, s)$ in inverse powers of s, which is usable in principle for any value of $s \geq 1$, but preferably for $s \geq 2$. The function $\exp\omega(\rho, s)$ has been tabulated by Morgan[2] for $1 \leq s \leq 2$.

It is evident that $\Omega(\rho)$ depends upon the choice of refractive index in the outer shell $a \leq r \leq 1$, being zero when $a = 1$. In principle it can be evaluated by numerical integration as soon as $P(r)$ is specified. If the index of refraction varies as any power of the radius in the outer shell, $\Omega(\rho)$ can be expressed[2] in terms of the function $\omega(\rho, s)$.

The function $\Theta(\rho)$ depends upon the exit angle $\theta(\kappa)$, which will be discussed below. If θ is proportional to α, a case already treated,[5,6] then $\Theta(\rho)$ is proportional to $\omega(\rho, r_0)$. Another interesting example is worked out in the last section of the present paper.

The general solution for the broad-beam Luneberg reflector of Fig. 1(b) is very similar to the preceding solution. If the index of refraction is specified according to (7) for $a \leq r \leq 1$, the condition that the marginal ray in the exit beam makes an angle ψ_0 with the negative x-axis leads to the restriction

$$\int_a^1 \frac{dr}{r[P^2(r) - 1]^{1/2}} \leq \frac{1}{4}\left[\pi + \sin^{-1}\frac{1}{r_0} - \psi_0\right]. \tag{14}$$

Since presumably $0 \leq \psi_0 \leq \pi$, the right side of (14) is never negative. The parametric equations which define the index of refraction in $0 \leq r \leq a$ are

$$n = (\rho^{1/2}/a)\exp\left[\tfrac{1}{2}\{\omega(\rho, r_0) + \omega(\rho, 1) - \Theta(\rho)\} - \Omega(\rho)\right],$$
$$r = \rho/n, \tag{15}$$

for $0 \leq \rho \leq 1$, where the functions on the right are defined, as before, by (11), (12), and (13).

In order to ascertain that (10) [or (15)] really does furnish a solution for (2) [or (3)], it is necessary to verify that $\rho(r)$ is a strictly increasing function of r; *i.e.*,

$$d\rho/dr > 0, \qquad 0 < \rho < 1. \tag{16}$$

A calculation similar to that given for the ordinary Luneberg lens[2] shows that the inequality (16) will be satisfied if $P(r)$ satisfies (8) [or (14)], and if $\theta(\kappa)$ is such that

$$\frac{d}{d\kappa}\left[\frac{\theta(\kappa) - \pi}{\kappa}\right] > 0, \qquad 0 < \kappa < 1. \tag{17}$$

The last inequality can be shown to be valid for the examples solved in the concluding section of this paper.

Energy Relationships

Let the desired antenna pattern be given by $F(\psi)$, for $0 \leq \psi \leq \psi_0$, where $F(\psi)$ represents the power flow per unit solid angle in the direction ψ, and all patterns are assumed to be independent of azimuth. Similarly let the feed pattern be given by $F_0(\alpha)$, for $0 \leq \alpha \leq \alpha_0$, where

$$\alpha = \sin^{-1}\frac{\kappa}{r_0}, \qquad \alpha_0 = \sin^{-1}\frac{1}{r_0}. \tag{18}$$

We assume that ψ is an increasing function of α and hence an increasing function of κ.

Neglecting any reflection losses[7] which may occur at discontinuities in index or at the surface of a Luneberg reflector, and allowing for arbitrary normalization of $F(\psi)$ and $F_0(\alpha)$, we have by conservation of energy,

$$\frac{\int_0^\alpha F_0(\alpha)\sin\alpha\,d\alpha}{\int_0^{\alpha_0} F_0(\alpha)\sin\alpha\,d\alpha} = \frac{\int_0^\psi F(\psi)\sin\psi\,d\psi}{\int_0^{\psi_0} F(\psi)\sin\psi\,d\psi}. \tag{19}$$

[3] Luneberg, "Mathematical Theory of Optics," Brown University, Providence, R. I., pp. 212–213; 1944.

[4] A. Fletcher, T. Murphy, and A. Young, "Solutions of two optical problems," *Proc. Roy. Soc. (London) A*, vol. 223, pp. 216–225; April 22, 1954.

[5] G. Toraldo di Francia, "A family of perfect configuration lenses of revolution," *Optica Acta*, vol. 1, pp. 157–163; February, 1955.

[6] R. Stettler, "Ueber radialsymmetrische optische Medien," *Optica Acta*, vol. 3, pp. 101–103; September, 1956.

[7] More precisely, we are neglecting *variations* in reflection loss as a function of angle. The effects of such variations may be computed in a straightforward manner for any given design, or they may be allowed for in advance if conditions warrant.

This equation gives ψ as an implicit function of α, which may be determined numerically in any case, and therefore gives ψ as a function of κ. From the geometry of Fig. 1 and Snell's law we obtain for the broad-beam Luneberg lens,

$$\theta(\kappa) = \psi(\kappa) + \phi(\kappa) = \psi(\kappa) + \sin^{-1}\kappa, \tag{20}$$

and for the broad-beam Luneberg reflector,

$$\theta(\kappa) = \psi(\kappa) + \phi(\kappa) - \pi = \psi(\kappa) + \sin^{-1}\kappa - \pi. \tag{21}$$

Examples

An interesting example in which the integrations can be performed analytically is given by a cosine-tapered feed and a uniform exit beam. Thus if we assume

$$F_0(\alpha) = \cos\alpha, \qquad 0 \leq \alpha \leq \alpha_0,$$
$$F(\psi) = 1, \qquad 0 \leq \psi \leq \psi_0, \tag{22}$$

(19) becomes

$$\frac{\sin^2\alpha}{\sin^2\alpha_0} = \frac{\sin^2\frac{1}{2}\psi}{\sin^2\frac{1}{2}\psi_0} = \kappa^2, \tag{23}$$

or

$$\psi = 2\sin^{-1}(\kappa\sin\tfrac{1}{2}\psi_0). \tag{24}$$

We then get for the broad-beam Luneberg lens,

$$\theta(\kappa) = 2\sin^{-1}(\kappa\sin\tfrac{1}{2}\psi_0) + \sin^{-1}\kappa, \tag{25}$$

and from (11) and (13),

$$\Theta(\rho) = 2\omega(\rho, \csc\tfrac{1}{2}\psi_0) + \omega(\rho, 1); \tag{26}$$

and finally, for the first of the parametric equations (10),

$$n = (1/a)\exp[\omega(\rho, r_0) - 2\omega(\rho, \csc\tfrac{1}{2}\psi_0) - \Omega(\rho)]. \tag{27}$$

Similarly, for the broad-beam Luneberg reflector,

$$\theta(\kappa) = 2\sin^{-1}(\kappa\sin\tfrac{1}{2}\psi_0) + \sin^{-1}\kappa - \pi, \tag{28}$$

and

$$\Theta(\rho) = 2\omega(\rho, \csc\tfrac{1}{2}\psi_0) - \omega(\rho, 1) + \log\rho; \tag{29}$$

and for the first of the parametric equations (15),

$$n = (1/a)\exp[\tfrac{1}{2}\omega(\rho, r_0) + \omega(\rho, 1) - \omega(\rho, \csc\tfrac{1}{2}\psi_0) - \Omega(\rho)]. \tag{30}$$

For the broad-beam Luneberg lens with the feed at the surface and no outer shell ($r_0 = 1$ and $a = 1$), (27) becomes

$$n = \exp[\omega(\rho, 1) - 2\omega(\rho, \csc\tfrac{1}{2}\psi_0)], \tag{31}$$

while for the broad-beam reflector with a source at infinity (parallel incident beam) and no outer shell, we get from (30),

$$n = \exp[\omega(\rho, 1) - \omega(\rho, \csc\tfrac{1}{2}\psi_0)]. \tag{32}$$

These are equivalent to the results given by Kay.

Acknowledgment

The author is indebted to A. F. Kay for sending him an account of his work[1] prior to publication.

Radiation Properties of a Thin Wire Loop Antenna Embedded in a Spherical Medium*

ORVAL R. CRUZAN†

Summary—**Formulas for certain radiation properties of a spherical antenna are derived theoretically. The antenna, which consists of a spherical medium, such as ferrite, with a thin wire loop embedded just below the surface in an equatorial plane, is driven by a slice generator. For the spherical medium, the permeability K_m and the dielectric constant K_e are assumed to be scalars and, in general, complex. The solutions are facilitated through the expansion of the fields in terms of characteristic orthogonal spherical vector wave functions. The properties for which formulas are derived are current distribution, input impedance, input power, radiated power, power loss in the spherical medium, and the efficiency of the antenna. For radiation resistance, not only the general case formula but also the formula for electrically small antennas is given, and the difference between these formulas, for media assumed lossless, is shown graphically.**

* Manuscript received by the PGAP, November 24, 1958; revised manuscript received, June 22, 1959.
† Diamond Ordnance Fuze Labs., Washington 25, D. C.

I. Introduction

ALTHOUGH a considerable amount of theoretical work has been published during the past few years about the properties of loaded antennas, most of it is concerned with electrically small antennas. Examples of some of this work for loaded antennas of

various sizes are available in the literature.[1–13] Since materials suitable for loading purposes, such as ferrites and dielectrics, have been and are being developed which have 1) low electrical conductivities, 2) small values of the loss tangents for both the permeabilities and the dielectric constants, and 3) values of the permeabilities and of the dielectric constants much greater than unity at VHF, it appears that the theory of loaded antennas of general size is desirable.

This paper gives[4] the theory and properties of an antenna which consists of a sphere of homogeneous isotropic material with a thin circular loop embedded just below the surface in an equatorial plane. The antenna, which is driven by a slice generator, is illustrated in Fig. 1. Another solution to this antenna problem has been given by Herman.[13] This is derived by first obtaining explicit expressions for the components of the electric and magnetic fields, using a magnetic vector potential. The theory of a general size antenna composed of a thin bicone embedded in a sphere of homogeneous isotropic material has been given by Schelkunoff.[12]

Of the loaded loop antenna, the various properties considered are current distribution, input impedance, input power, radiated power, power loss in the loading material, and the efficiency of the antenna. Knowing the current distribution, it is possible to determine the remaining properties of the antenna. Consequently, the first part of the paper is devoted to deriving the integral

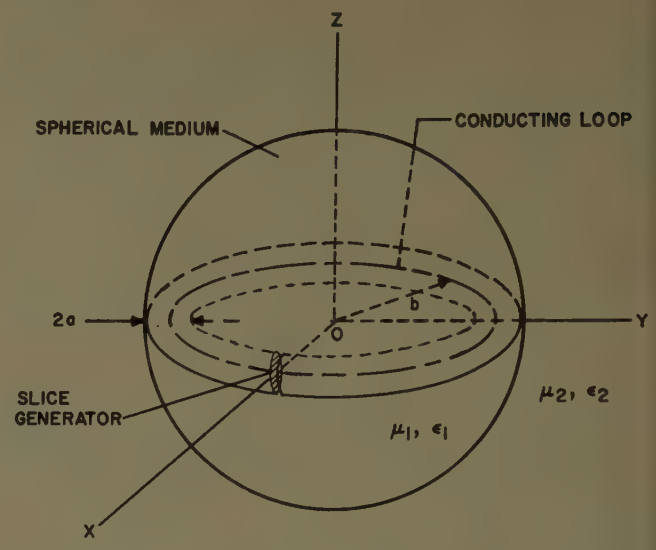


Fig. 1—Circular loop antenna embedded in a spherical medium.

equation for the determination of the current distribution when the input voltage is known. In deriving the integral equation, the field, which is expressed as a function of a retarded Hertz vector potential, is reduced, first, for the case of a thin circular loop. The field is then expanded in terms of characteristic orthogonal spherical vector wave functions through such an expansion of the retarded Hertz vector potential. The tangential component of the total electric field at the surface of the sphere and of the loop is obtained through use of the boundary conditions pertaining to the tangential components. Upon relating the tangential component of the electric field to the voltage of a slice generator by means of a delta function, the integral equation is obtained. It will be observed that the method employed can be readily extended to other arrangements of loop and sphere. In such cases the formulas obtained will be more complicated, since more separate boundaries are taken into consideration.

Since the expressions for the various properties are in the form of infinite series, adequate evaluations involving ranges of values of the various parameters would require the use of a high-speed digital computer. Fig. 3, which is an example of partial results obtained using a desk calculator, illustrates how the radiation resistance given by the formula for the general size antenna, using loading material assumed lossless, contrasts to that given by the formula for the electrically small antenna.

II. General Equations[14]

The electric and the magnetic fields of a radiating source embedded in an infinite homogeneous and isotropic medium are given, respectively, by

[1] L. Page, "The magnetic antenna," *Phys. Rev.*, vol. 69, pp. 645–648; June, 1946. (Electrically large cylindrical rod and circular coil.)

[2] O. R. Cruzan, "VHF ferrite antenna, I. Radiation properties," *1957 PGMIL Conv. Rec.*, pp. 169–175. (Mainly electrically small spherical core.)

[3] O. R. Cruzan, "VHF Ferrite Antenna Radiation Properties," Diamond Ordnance Fuze Labs., Washington 25, D. C., Tech. Rept. No. TR-516; August 15, 1957.

[4] O. R. Cruzan, "Radiation Properties of a Spherical Ferrite Antenna," Diamond Ordnance Fuze Labs., Washington 25, D. C., Tech. Rept. No. TR-387; October 15, 1956. (General size spherical core.)

[5] H. A. Dropkin, E. Metzger, and J. C. Cacheris, "VHF ferrite antenna, II. Radiation measurements," *1957 PGMIL Conv. Rec.*, pp. 175–182. (Ferrite rod cores.)

[6] V. H. Rumsey and W. L. Weeks, "Impedance of Ferrite Loop Antennas," Elec. Engrg. Res. Lab., Univ. of Illinois, Urbana, Tech. Rept. No. 13; October 15, 1956. (Electrically small antennas.)

[7] W. L. Weeks, "Input Impedance of a Spherical Ferrite Antenna with a Latitudinal Current," Elec. Engrg. Res. Lab., Univ. of Illinois, Urbana, Tech. Rept. No. 6; August 20, 1955. (Electrically small antennas.)

[8] W. L. Weeks, "On the Estimation of Ferrite Loop Antenna Impedance," Elec. Engrg. Res. Lab., Univ. of Illinois, Urbana, Tech. Rept. No. 17; April 10, 1957. (Electrically small antennas.)

[9] J. R. Wait, "The receiving loop with a hollow prolate spheroidal core," *Canad. J. Tech.*, vol. 31, pp. 132–137; June, 1953. (Electrically small antenna.)

[10] J. R. Wait, "Receiving properties of a wire loop with a spheroidal core," *Canad. J. Tech.*, vol. 31, pp. 9–14; January, 1953.

[11] R. E. Burgess, "Iron-cored loop receiving aerial," *Wireless Engr.*, vol. 23, pp. 172–178; June, 1946. (Spheroidal core with coil.)

[12] S. A. Schelkunoff, "Advanced Antenna Theory," John Wiley and Sons, Inc., New York, N. Y., pp. 65–68; 1952. (Thin bicone embedded in a sphere.)

[13] J. Herman, "Thin Wire Loop and Thin Biconical Antennas in Finite Spherical Media," Diamond Ordnance Fuze Labs., Washington 25, D. C., Tech. Rept. No. TR-462; May 1, 1957. (General size spherical core.)

[14] J. A. Stratton, "Electromagnetic Theory," McGraw-Hill Book Co., Inc., New York, N. Y., sec. 8.3, pp. 430–431, and ch. 7, pp. 392–423; 1941.

$$E = \nabla\nabla\cdot\pi' + k^2\pi' = \nabla\times\nabla\times\pi', \tag{1}$$

the second form being for the field outside the region occupied by the source, and

$$H = -i\omega\epsilon\nabla\times\pi', \tag{2}$$

where the time factor is of the form $(\exp -i\omega t)$, $k^2=\mu\epsilon\omega^2$, μ is the magnetic inductive capacity of the medium, ϵ is the electric inductive capacity, and ω is the radian measure of the frequency. For a radiating current source, the retarded Hertz vector potential π' is given by

$$\pi' = \frac{1}{4\pi\omega\epsilon}\int_V iJ_0\frac{e^{ikR}}{R'}\,dv, \tag{3}$$

where J_0 is the vector current density, V is the volume occupied by the current, and R' is the distance from a current element to a field point. Inside the region occupied by the current, the vector potential as given by (3) satisfies the inhomogeneous vector wave equation

$$\nabla\times\nabla\times\pi' - \nabla\nabla\cdot\pi' - k^2\pi' = \frac{i}{\omega\epsilon}J_0. \tag{4a}$$

Outside that region, the homogeneous vector wave equation

$$\Delta\times\Delta\times\pi' - \nabla\nabla\cdot\pi' - k^2\pi' = 0 \tag{4b}$$

is satisfied. Because (4b) permits the potential to be expanded in terms of characteristic vector wave functions, the potential, accordingly, will be obtained for field points external to the region occupied by the current. The characteristic orthogonal spherical vector wave functions used are

$$m_{\substack{e\\o}mn} = \mp\frac{m}{\sin\theta}z_n(kr)P_n^m(\cos\theta)\begin{matrix}\sin\\\cos\end{matrix}m\phi i_\theta - z_n(kr)\frac{\partial}{\partial\theta}P_n^m(\cos\theta)\begin{matrix}\cos\\\sin\end{matrix}m\phi i_\phi, \tag{5}$$

$$n_{\substack{e\\o}mn} = \frac{n(n+1)}{kr}z_n(kr)P_n^m(\cos\theta)\begin{matrix}\cos\\\sin\end{matrix}m\phi i_r + \frac{1}{kr}\frac{\partial}{\partial r}[rz_n(kr)]\frac{\partial}{\partial\theta}P_n^m(\cos\theta)\begin{matrix}\cos\\\sin\end{matrix}m\theta i_\theta \mp \frac{m}{kr\sin\theta}\frac{\partial}{\partial r}[rz_n(kr)]P_n^m(\cos\theta)\begin{matrix}\sin\\\cos\end{matrix}m\phi i_\phi, \tag{6}$$

where $z_n(kr)=(\pi/2kr)^{1/2}Z_{n+1/2}(kr)$ stands for any spherical Bessel function with $Z_{n+1/2}(kr)$ being the corresponding cylindrical Bessel function of half order, and $P_n^m(\cos\theta)$ is the associated Legendre polynomial of the first kind.

For a current-bearing thin circular loop having a circular cross section of radius a, it may be shown that

$$\pi' \simeq \pi = \frac{b}{4\pi\omega\epsilon}\int_0^{2\pi} iI(\phi')\frac{e^{ikR}}{R}i_{\phi'}\,d\phi', \tag{7}$$

where b is the mean radius of the loop, $I(\phi')$ is the total current on the conductor surface, $i_{\phi'}$ is a unit vector having the same direction of J_0, and R is the distance from the center of the conductor forming the loop. The distances, angles and vectors involved are illustrated in Figs. 2(a) and 2(b).

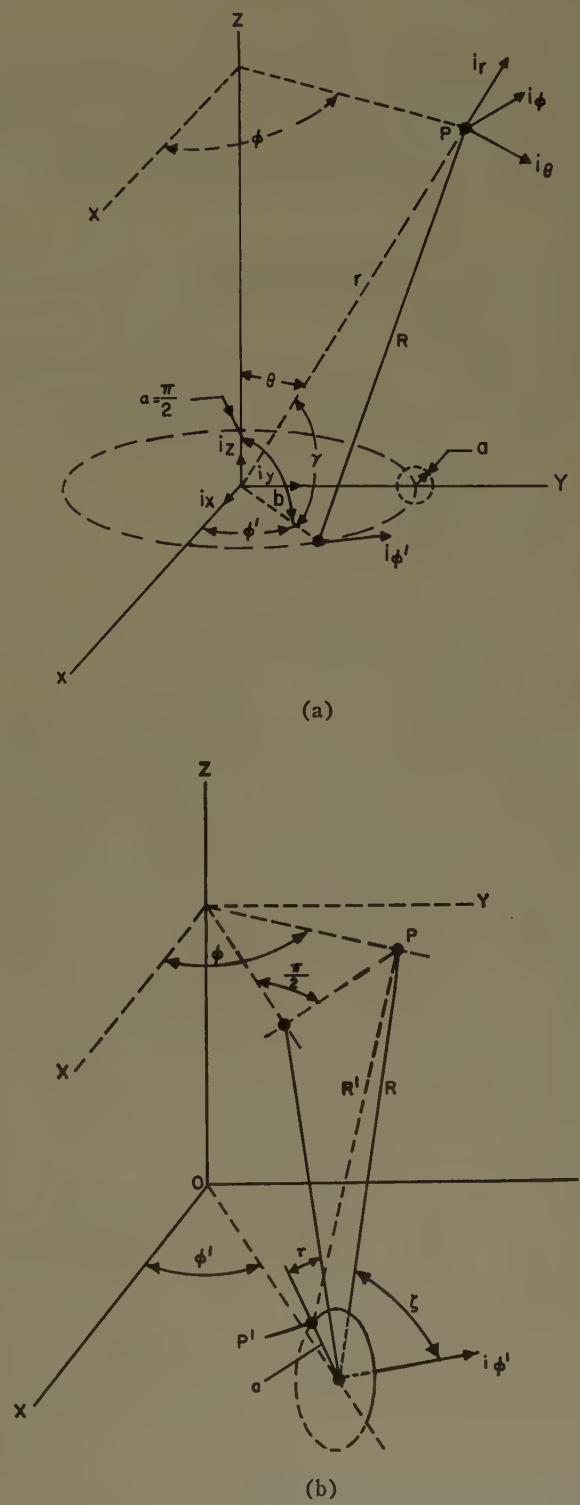


Fig. 2—(a) Unit vectors, angles and distances associated with a thin circular loop. (b) Angles and distances associated with a cross section of conductor.

The Green's function e^{ikR}/R is given by the addition theorem

$$\frac{e^{ikR}}{R} = ik\sum_{p=0}^{\infty}(2p+1)P_p(\cos\gamma)j_p(kr)h_p^{(1)}(kb), \quad r<b \tag{8a}$$

$$= ik\sum_{p=0}^{\infty}(2p+1)P_p(\cos\gamma)j_p(kb)h_p^{(1)}(kr), \quad r>b, \tag{8b}$$

where γ is the angle between b and r (the vector distance from the center of the loop to the field point P), $j_p(kr)$ is a spherical Bessel function of the first kind, and $h_p^{(1)}(kr)$ is a spherical Hankel function of the first kind. The expansion of the Legendre polynomial $P_p(\cos\gamma)$ is

$$P_p(\cos\gamma) = 2\sum_{q=0}^{p}\frac{1}{1+\delta}\frac{(p-q)!}{(p+q)!}P_p^q(\cos\alpha)P_p^q(\cos\theta)\cos q(\phi'-\phi), \tag{9}$$

where $\delta=0$, $q\neq 0$; $\delta=1$, $q=0$, and $\alpha=\pi/2$. The various distances and angles involved in (8) and (9) are illustrated in Fig. 2(a). As can be determined, the unit vector $i_{\phi'}$ is given by

$$i_{\phi'} = -i_r\sin\theta\sin(\phi'-\phi) - i_\theta\cos\theta\sin(\phi'-\phi) + i_\phi\cos(\phi'-\phi). \tag{10}$$

III. Expansion of the Primary Field Equations in Vector Wave Functions

Using the expression for the retarded Hertz vector potential for the case of a thin circular loop, it is possible to obtain explicit expressions for the electric and the magnetic fields in terms of vector wave functions. This is done by making such an expansion of the retarded Hertz vector potential π as given by (7). Since $I(\phi')$ is independent of the coordinates of the field, then to expand π, it is sufficient to expand only the remaining part of the integrand of (7) in terms of characteristic orthogonal spherical vector wave functions.[14]

Thus for $r>b$, in an infinite medium having a permeability K_m and a dielectric constant K_e,

$$\boldsymbol{E} = \frac{k^2\beta}{2\pi\omega\epsilon}\int_0^{2\pi}I(\phi')\sum_{n=0}^{\infty}\sum_{m=0}^{n}\cdot[a_{mn}^{i}(\boldsymbol{m}_{omn}^{(3)}\sin m\phi' + \boldsymbol{m}_{emn}^{(3)}\cos m\phi') + b_{mn}^{i}(\boldsymbol{n}_{emn}^{(3)}\sin m\phi' - \boldsymbol{n}_{omn}^{(3)}\cos m\phi')]d\phi'$$

$$\boldsymbol{H} = \frac{-ik\beta}{2\pi}\int_0^{2\pi}I(\phi')\sum_{n=0}^{\infty}\sum_{m=0}^{n}\cdot[a_{mn}^{i}(\boldsymbol{n}_{omn}^{(3)}\sin m\phi' + \boldsymbol{n}_{emn}^{(3)}\cos m\phi') + b_{mn}^{i}(\boldsymbol{m}_{emn}^{(3)}\sin m\phi' - \boldsymbol{m}_{omn}^{(3)}\cos m\phi')]d\phi', \tag{11}$$

where $\beta=kb$; the superscript (3) means that in the vector wave functions $h_n^{(1)}(kr)$ has been substituted for $z_n(kr)$. The coefficients are

$$a_{mn}^{i} = \frac{-(2n+1)}{n(n+1)(1+\delta)}\frac{(n-m)!}{(n+m)!}j_n(\beta)P_n^{m+1}(0)$$

$$b_{mn}^{i} = \frac{m(2n+1)}{n(n+1)(1+\delta)}\frac{(n-m)!}{(n+m)!}\frac{[\beta j_n(\beta)]'}{\beta}P_n^m(0), \tag{12}$$

where $\delta=0$, $m\neq 0$; $\delta=1$, $m=0$. The prime on the quantity within brackets, *e.g.*, $[xh_n^{(1)}(x)]'$ means differentiation with respect to the argument x.

IV. The Fields of a Two-Media Region

If the field in an infinite medium is known, then the fields for a two-media region can be obtained. This two-media region, illustrated in Fig. 1, consists of a sphere of material, such as ferrite, for which $r=c\geq(b+a)$, and external free space. The fields of the two media are obtained by relating through the tangential boundary conditions,

$$i_r\times(\boldsymbol{E}_2-\boldsymbol{E}_1)=0 \qquad i_r\times(\boldsymbol{H}_1-\boldsymbol{H}_1)=0, \tag{13}$$

where i_r is the unit radial vector. The field within the sphere consists of the primary and the secondary fields, that is

$$\boldsymbol{E}_1=\boldsymbol{E}_i+\boldsymbol{E}_r, \qquad \boldsymbol{H}_1=\boldsymbol{H}_i+\boldsymbol{H}_r, \tag{14}$$

where the subscript i refers to the primary field and the subscript r refers to the secondary field. For the transmitted field,

$$\boldsymbol{E}_2=\boldsymbol{E}_t, \qquad \boldsymbol{H}_2=\boldsymbol{H}_t. \tag{15}$$

The primary field is obtained from (11) by attaching the subscript 1 to the parameters to indicate the spherical medium in which the field is situated; thus

$$\boldsymbol{E}_i = \frac{k_1^2\beta_1}{2\pi\omega\epsilon_1}\int_0^{2\pi}I(\phi')\sum_{n=0}^{\infty}\sum_{m=0}^{n}\cdot[a_{mn}^{i}(\boldsymbol{m}_{omn}^{(3)}\sin m\phi' + \boldsymbol{m}_{emn}^{(3)}\cos m\phi') + b_{mn}^{i}(\boldsymbol{n}_{emn}^{(3)}\sin m\phi' - \boldsymbol{n}_{omn}^{(3)}\cos m\phi')]d\phi',$$

$$\boldsymbol{H}_i = \frac{-ik_1\beta_1}{2\pi}\int_0^{2\pi}I(\phi')\sum_{n=0}^{\infty}\sum_{m=0}^{n}\cdot[a_{mn}^{i}(\boldsymbol{n}_{omn}^{(3)}\sin m\phi' + \boldsymbol{n}_{emn}^{(3)}\cos m\phi') + b_{mn}^{i}(\boldsymbol{m}_{emn}^{(3)}\sin m\phi' - \boldsymbol{m}_{omn}^{(3)}\cos m\phi')]d\phi'. \tag{16}$$

For the secondary field in the spherical medium:

$$\boldsymbol{E}_r = \frac{k_1^2\beta_1}{2\pi\omega\epsilon_1}\int_0^{2\pi}I(\phi')\sum_{n=0}^{\infty}\sum_{m=0}^{n}\cdot[a_{mn}^{r}(\boldsymbol{m}_{omn}^{(1)}\sin m\phi' + \boldsymbol{m}_{emn}^{(1)}\cos m\phi') + b_{mn}^{r}(\boldsymbol{n}_{emn}^{(1)}\sin m\phi' - \boldsymbol{n}_{omn}^{(1)}\cos m\phi')]d\phi',$$

$$H_r = \frac{-ik_1\beta_1}{2\pi}\int_0^{2\pi} I(\phi')\sum_{n=0}^{\infty}\sum_{m=0}^{n} \cdot[a_{mn}^{r}(n_{omn}^{(1)}\sin m\phi' + n_{emn}^{(1)}\cos m\phi') + b_{mn}^{r}(m_{emn}^{(1)}\sin m\phi' - m_{omn}^{(1)}\cos m\phi')]d\phi', \quad (17)$$

where the superscript (1) is used to indicate that the spherical Hankel function $h_n^{(1)}(x)$ has been replaced by the spherical Bessel function $j_n(x)$, since the radiation must be finite at the origin. For the transmitted field in free space:

$$E_t = \frac{k_2^2\beta_2}{2\pi\omega\epsilon_2}\int_0^{2\pi} I(\phi')\sum_{n=0}^{\infty}\sum_{m=0}^{n} \cdot[a_{mn}^{t}(m_{omn}^{(3)}\sin m\phi' + m_{emn}^{(3)}\cos m\phi') + b_{mn}^{t}(n_{emn}^{(3)}\sin m\phi' - n_{omn}^{(3)}\cos m\phi')]d\phi'$$

$$H_t = \frac{-ik_2\beta_2}{2\pi}\int_0^{2\pi} I(\phi')\sum_{n=0}^{\infty}\sum_{m=0}^{n} \cdot[a_{mn}^{t}(n_{omn}^{(3)}\sin m\phi' + n_{emn}^{(3)}\cos m\phi') + b_{mn}^{t}(m_{emn}^{(3)}\sin m\phi' - m_{omn}^{(3)}\cos m\phi')]d\phi'. \quad (18)$$

It is to be noted that although the same notation is used for the vector wave functions in (16) and (18), the argument of the spherical Hankel functions in (16) is k_1r and in (18), k_2r. In the three sets of equations, (16)–(18), $k_1=\omega(\mu_1\epsilon_1)^{1/2}$, $k_2=\omega(\mu_2\epsilon_2)^{1/2}$, $\beta_1=k_1b$, and $\beta_2=k_2b$.

The application of the boundary conditions (13) to (16)–(18) for $r=c\rightarrow(b+a)$ results in the two sets of simultaneous equations:

$$a_{mn}^{r}K_m(K_mK_e)^{1/2}j_n(\xi_1) - a_{mn}^{t}h_n^{(1)}(\xi_2) = -a_{mn}^{i}K_m(K_mK_e)^{1/2}h_n^{(1)}(\xi_1)$$

$$a_{mn}^{r}(K_mK_e)^{1/2}[\xi_1j_n(\xi_1)]' - a_{mn}^{t}[\xi_2h_n^{(1)}(\xi_2)]' = -a_{mn}^{i}(K_mK_e)^{1/2}[\xi_1h_n^{(1)}(\xi_1)]', \quad (19)$$

and

$$b_{mn}^{r}K_mK_ej_n(\xi_1) - b_{mn}^{t}h_n^{(1)}(\xi_2) = -b_{mn}^{i}K_mK_eh_n^{(1)}(\xi_1)$$

$$b_{mn}^{r}K_m[\xi_1j_n(\xi_1)]' - b_{mn}^{t}[\xi_2h_n^{(1)}(\xi_2)]' = -b_{mn}^{i}K_m[\xi_1h_n^{(1)}(\xi_1)]', \quad (20)$$

where $K_m=\mu_1/\mu_2$; $K_e=\epsilon_1/\epsilon_2$; $\xi_1=k_1(b+a)$ and $\xi_2=k_2(b+a)$. Since for determining the current distribution and the input impedance, the field within the sphere is of primary interest, then from (19) and (20), respectively:

$$a_{mn}^{r} = -a_{mn}^{i}\frac{K_mh_n^{(1)}(\xi_1)[\xi_2h_n^{(1)}(\xi_2)]' - h_n^{(1)}(\xi_2)[\xi_1h_n^{(1)}(\xi_1)]'}{K_mj_n(\xi_1)[\xi_2h_n^{(1)}(\xi_2)]' - h_n^{(1)}(\xi_2)[\xi_1j_n(\xi_1)]'}, \quad (21)$$

and

$$b_{mn}^{r} = -b_{mn}^{i}\frac{K_eh_n^{(1)}(\xi_1)[\xi_2h_n^{(1)}(\xi_2)]' - h_n^{(1)}(\xi_2)[\xi_1h_n^{(1)}(\xi_1)]'}{K_ej_n(\xi_1)[\xi_2h_n^{(1)}(\xi_2)]' - h_n^{(1)}(\xi_2)[\xi_1j_n(\xi_1)]'}. \quad (22)$$

The total electric field within the sphere at $r=(b+a)$, due to the current $I(\phi')$, is the sum of the electric fields as given by (16) and (17).

V. Current Distribution and Input Impedance

To obtain the current distribution and the input impedance from the electric field, the tangential component of the total electric field at the surface, $r=(b+a)$, of the conducting loop is needed. Let E^e represent this component, then

$$E^e = -i\left(\frac{\mu_2}{\epsilon_2}\right)^{1/2}\frac{K_m}{\lambda}\frac{\beta_1}{\xi_1}\int_0^{2\pi} I(\phi')\cdot\sum_{n=0}^{\infty}\sum_{m=0}^{n} A_{mn}\cos m(\phi'-\phi)d\phi', \quad (23)$$

where

$$A_{mn} = \frac{1}{1+\delta}\left(a_{mn}a_n + \frac{m^2K_e}{\beta_1\xi_1}b_{mn}b_n\right) \quad (24)$$

with

$$a_{mn} = \frac{2n+1}{n(n+1)}\frac{(n-m)!}{(n+m)!}[P_n^{m+1}(0)]^2 \quad (25)$$

$$a_n = \frac{j_n(\beta_1)h_n^{(1)}(\xi_2)}{K_mj_n(\xi_1)[\xi_2h_n^{(1)}(\xi_2)]' - h_n^{(1)}(\xi_2)[\xi_1j_n(\xi_1)]'} \quad (26)$$

$$b_{mn} = \frac{2n+1}{n(n+1)}\frac{(n-m)!}{(n+m)!}[P_n^{m}(0)]^2 \quad (27)$$

$$b_n = \frac{[\beta_1j_n(\beta_1)]'[\xi_2h_n^{(1)}(\xi_2)]'}{K_ej_n(\beta_1)[\xi_2h_n^{(1)}(\xi_2)]' - h_n^{(1)}(\xi_2)[\xi_1j_n(\xi_1)]'}. \quad (28)$$

For a body which is a perfect conductor, as is assumed for the loop, the tangential component of the electric field at the surface is zero. If the loop contains a gap,[15–17] then the only part of the tangential component of the electric field not zero is in the gap. Thus there is obtained from (23)

[15] C. Flammer, "The Prolate Spheroidal Monopole Antenna," Stanford Res. Inst., Menlo Park, Calif., Tech. Rept. No. 22; June, 1957.

[16] L. Infeld, "The influence of the width of the gap upon the theory of antennas," *Quart. Appl. Math.*, vol. 5, pp. 113–132; July, 1947.

[17] J. A. Stratton and L. J. Chu, "Steady state solutions of electromagnetic field problems," *J. Appl. Phys.*, vol. 12, pp. 230–248; March, 1941.

$$E^e = 0 \qquad \text{(outside the gap)}$$

$$F^i = E^e = -i\left(\frac{\mu_2}{\epsilon_2}\right)^{1/2} \frac{K_m}{\lambda} \frac{\beta_1}{\xi_1} \int_0^{2\pi} I(\phi') \cdot \sum_{n=0}^{\infty} \sum_{m=0}^{n} A_{mn} \cos m(\phi' - \phi) d\phi' \quad \text{(inside the gap)}. \tag{29}$$

If the driving voltage for the antenna is produced in the gap, *i.e.*, by a slice generator, then

$$bF^i = -V^i \delta(\phi), \tag{30}$$

where V^i is the input voltage at $\phi = 0$ and $\delta(\phi)$ is a delta function defined as

$$\delta(\phi) = \lim_{\nu \to 0} \frac{b}{\sqrt{2\pi}} \frac{e^{-b^2\phi^2/2\nu^2}}{\nu}. \tag{31}$$

Eq. (30) merely indicates that both sides when used as integrands give the same results. The substitution of the value of F^i from (29) into (30) gives

$$V^i \delta(\phi) = \frac{i}{2\pi}\left(\frac{\mu_2}{\epsilon_2}\right)^{1/2}\left(\frac{K_m}{K_e}\right)^{1/2} \frac{\beta_1^2}{\xi_1} \int_0^{2\pi} I(\phi') \cdot \sum_{n=0}^{\infty} \sum_{m=0}^{n} A_{mn} \cos m(\phi' - \phi) d\phi'. \tag{32}$$

Assume that $I(\phi')$ can be represented by the Fourier expansion,

$$I(\phi') = \sum_{p=0}^{\infty} \frac{1}{1+\delta} (A_p \cos p\phi' + B_p \sin p\phi'), \tag{33}$$

where $\delta = 0$, $p \neq 0$; $\delta = 1$, $p = 0$. Eq. (32) now becomes

$$V^i \delta(\phi) = \frac{1}{2}\left(\frac{\mu_2}{\epsilon_2}\right)^{1/2}\left(\frac{K_m}{K_e}\right)^{1/2} \frac{\beta_1^2}{\xi_1} \cdot \sum_{n=0}^{\infty} \sum_{m=0}^{n} A_m (A_m \cos m\phi + B_m \sin m\phi). \tag{34}$$

The coefficients A_m and B_m are determined by making, respectively, the Fourier cosine and sine transforms. These transforms of $V^i\delta(\phi)$ are

$$\lim_{\nu \to 0} \int_{-\pi}^{\pi} V^i \delta(\phi) \cos p\phi d\phi \to \lim_{\nu \to 0} V^i e^{-p^2\nu^2/2b^2} \to V^i$$

and

$$\lim_{\nu \to 0} \int_{-\pi}^{\pi} V^i \delta(\phi) \sin p\phi d\phi = 0. \tag{35}$$

Transforms (35) give

$$A_m = \frac{V^i}{i\frac{\pi}{2}\left(\frac{\mu_2}{\epsilon_2}\right)^{1/2}\left(\frac{K_m}{K_e}\right)^{1/2} \frac{\beta_1^2}{\xi_1}(1+\delta) \sum_{n=0}^{\infty} A_{mn}}$$

$$B_m = 0. \tag{36}$$

The substitution of these values of A_m and B_m into (33) gives

$$I(\phi') = \frac{-2V^i i}{\pi\left(\frac{\mu_2}{\epsilon_2}\right)^{1/2}\left(\frac{K_m}{K_e}\right)^{1/2} \frac{\beta_1^2}{\xi_1}} \sum_{m=0}^{\infty} \frac{\cos m\phi'}{(1+\delta)^2 \alpha_m}, \tag{37}$$

where

$$\alpha_m = \sum_{n=0}^{\infty} A_{mn} = \frac{1}{1+\delta} \sum_{n=0}^{\infty} \left(a_{mn}a_n + \frac{m^2 K_e}{\beta_1 \xi_1} b_{mn} b_n\right). \tag{38}$$

To obtain the input impedance, let $\phi' = 0$, then

$$I(0) = \frac{-2V^i i}{\pi\left(\frac{\mu_2}{\epsilon_2}\right)^{1/2}\left(\frac{K_m}{K_e}\right)^{1/2} \frac{\beta_1^2}{\xi_1}} \sum_{m=0}^{\infty} \frac{1}{(1+\delta)^2 \alpha_m}. \tag{39}$$

Since

$$Z^i = \frac{V^i}{I(0)}, \tag{40}$$

then the use of the value of $V^i/I(0)$, obtained from (39), gives

$$Z^i = \frac{i}{2} \pi \left(\frac{\mu_2}{\epsilon_2}\right)^{1/2}\left(\frac{K_m}{K_e}\right)^{1/2} \frac{\beta_1^2}{\xi_1}\left[\sum_{m=0}^{\infty} \frac{1}{(1+\delta)^2 \alpha_m}\right]^{-1}. \tag{41}$$

VI. The Input Power

The input power of a loaded antenna is composed of three parts: 1) the radiated power W_r, 2) the power loss W_f due to the complex parameters K_m and K_e of the loading medium, and 3) the power loss W_c due to the ohmic resistance of the conductor forming the loop. The power loss W_c due to the ohmic resistance of the conductor may be obtained to a first approximation by considering the skin effect. In thus determining this loss, it is assumed that the field distribution, obtained using a perfect conductor, is modified, at most, only a negligible amount. For copper, the resistance ρ, in ohms per meter, due to the skin effect, is[18]

$$\rho = \frac{1}{2\pi a} \frac{0.0143}{\sqrt{10\lambda}}, \tag{42}$$

where a is the radius of the conductor in meters, and λ is the wavelength of the radiation in meters as measured in free space. For a differential of length, the conductor power loss is

$$dW_c = \tfrac{1}{2}\rho |I|^2 dl, \tag{43}$$

where for a circular loop,

$$dl = bd\phi'. \tag{44}$$

Since ρ and b are fixed quantities for a given loop and a given frequency but I is a function of ϕ', then the total power loss is

[18] S. A. Schelkunoff and H. T. Friis, "Antennas: Theory and Practice," John Wiley and Sons, New York, N. Y., sec. 11.8, p. 339; 1952.

$$W_c = \tfrac{1}{2}\rho b \int_0^{2\pi} |I|^2 d\phi'. \tag{45}$$

The substitution of the values of ρ and I from (42) and (37), respectively, into (45), gives

$$W_c = \frac{b}{a}\,\frac{0.0143}{\sqrt{10\lambda}}\,V^{i^2}|A|^2 \sum_{m=0}^{\infty} \frac{1}{(1+\delta)^3 |\alpha_m|^2}, \tag{46}$$

where

$$A = \frac{1}{\pi\left(\frac{\mu_2}{\epsilon_2}\right)^{1/2}\left(\frac{K_m}{K_e}\right)^{1/2}\frac{\beta_1^2}{\xi_1}}, \tag{47}$$

as the power loss due to the skin effect of the conductor forming the loop.

The other parts of the input power W_r and W_f can be obtained in a combined form by using the average power formula. The separate determination of each is given in Sections VII and VIII. The average power formula is

$$W_{rf} = \tfrac{1}{2}\,\mathrm{Re}\,[V^i I^*], \tag{48}$$

where W_{rf} is the combine power due to radiation in free space and to losses in the loading medium, V^i is the input voltage, and I^* is the complex conjugate of the input current. The substitution of the complex conjugate of the input current, obtained from (39), into (48) gives for the input power W_{rf},

$$W_{rf} = V^{i^2}\,\mathrm{Re}\left[iA^* \sum_{m=0}^{\infty} \frac{1}{(1+\delta)^2 \alpha_m^*}\right], \tag{49}$$

where A is given by (47). The total input power consists of the sum of (49) and (46), that is,

$$W^i = W_{rf} + W_c. \tag{50}$$

VII. The Radiated Power

The radiated power is obtained by integrating the normal component of the complex Poynting vector S^* over the surface of an infinite sphere, that is

$$W_r = \lim_{r\to\infty} \int_0^{2\pi}\int_0^{\pi} S^* \cdot i_r r^2 \sin\theta\, d\theta\, d\phi, \tag{51}$$

where

$$S^* = \tfrac{1}{2} E_t \times H_t^*. \tag{52}$$

From (33) and (36) it is seen that $I(\phi')$ is of the form

$$I(\phi') = \sum_{m=0}^{\infty} \frac{A_m}{1+\delta}\cos m\phi'. \tag{53}$$

If this expression of $I(\phi')$ is used in (18), the transmitted field takes the form

$$E_t = \frac{k_2^2\beta_2}{2\omega\epsilon_2} \sum_{n=0}^{\infty}\sum_{m=0}^{n} A_m(a_{mn}^t \mathbf{m}_{emn}^{(3)} - b_{mn}^t \mathbf{n}_{omn}^{(3)})$$

$$H_t = \frac{-ik_2\beta_2}{2} \sum_{q=0}^{\infty}\sum_{p=0}^{q} A_p(a_{pq}^t \mathbf{n}_{epq}^{(3)} - b_{pq}^t \mathbf{n}_{opq}^{(3)}),$$

where a_{mn}^t, a_{pq}^t, and b_{mn}^t, b_{pq}^t are obtained by solving (19) and (20), respectively. The complex Poynting vector becomes

$$S^* = \frac{ik_2^3\beta_2^2}{8\omega\epsilon_2} \sum_{n=0}^{\infty}\sum_{m=0}^{n}\sum_{q=0}^{\infty}\sum_{p=0}^{q} \cdot A_m A_p^* [a_{mn}^t a_{pq}^{t*} \mathbf{m}_{emn}^{(3)} \times \mathbf{n}_{epq}^{(3)*} + b_{mn}^t b_{pq}^{t*} \mathbf{n}_{omn}^{(3)} \times \mathbf{m}_{opq}^{(3)*} - a_{mn}^t b_{pq}^{t*} \mathbf{m}_{emn}^{(3)} \times \mathbf{m}_{opq}^{(3)*} - b_{mn}^t a_{pq}^{t*} \mathbf{n}_{omn}^{(3)} \times \mathbf{n}_{epq}^{(3)*}]. \tag{55}$$

The integral of the normal component of the complex Poynting vector over the surface of an infinite sphere, (51), gives for the radiated power

$$W_r = \frac{\epsilon_2\beta_2^2 V^{i^2}}{16\pi^5}\frac{K_m}{K_e}\left(\frac{b}{\lambda}\right)^4 \sum_{n=0}^{\infty}\sum_{m=0}^{n} \frac{2n+1}{n(n+1)}\,\frac{(n-m)!}{(n+m)!}\,\frac{1}{(1+\delta)^3} \cdot \left\{\left|\frac{K_m}{K_e}\right| [P_n^{m+1}(0)]^2 |a_n^t|^2 + \frac{m^2}{|\beta_1|^2}[P_n^m(0)]^2 |b_n^t|^2\right\}, \tag{56}$$

where

$$a_n^t = \frac{j_n(\beta_1)}{K_m j_n(\xi_1)[\xi_2 h_n^{(1)}(\xi_2)]' - h_n^{(1)}(\xi_2)[\xi_1 j_n(\xi_1)]'}$$

$$b_n^t = \frac{[\beta_1 j_n(\beta_1)]'}{K_e j_n(\xi_1)[\xi_2 h_n^{(1)}(\xi_2)]' - h_n^{(1)}(\xi_2)[\xi_1 j_n(\xi_1)]'}. \tag{57}$$

VIII. Power Loss in the Medium and Efficiency of the Antenna

The power loss W_f in the medium due to the complex parameters K_m and K_e is obtained by subtracting the radiated power, as given by (56), from the part of the input power, as given by (49). Thus the power loss in the medium is

$$W_f = W_{rf} - W_r. \tag{57}$$

The efficiency of a radiating antenna is defined as the ratio of the radiated power to the total input power. Thus η, the efficiency, is

$$\eta = \frac{W_r}{W_r + W_f + W_c}, \tag{58}$$

where W_r, the radiated power; W_f, the power loss in the medium; and W_c, the power loss due to the ohmic resistance of the loop, are given by (56), (57), and (46), respectively.

IX. Special Cases and Some Results

A. Thin Circular Loop in Free Space

The case of a thin circular loop in free space is obtained if $K_m = K_e = 1$. This gives, from (37), for the current distribution:

$$I(\phi') = \frac{V^i}{240\pi^4(b/\lambda)^2} \sum_{m=0}^{\infty} \frac{\cos m\phi'}{(1+\delta)^2 \alpha_m'}. \tag{59}$$

Eq. (41), for the input impedance, becomes

$$Z^i = 240\pi^4(b/\lambda)^2\left[\sum_{m=0}^{\infty}\frac{1}{(1+\delta)^2\alpha_m'}\right]^{-1}, \tag{60}$$

where

$$\alpha_m' = \sum_{n=0}^{\infty}\left(a_{mn}a_n' + \frac{m^2}{\beta\xi}b_{mn}b_n'\right) \tag{61}$$

$$a_{mn} = \frac{2n+1}{n(n+1)}\frac{(n-m)!}{(n+m)!}[P_n^{m+1}(0)]^2 \tag{62}$$

$$a_n' = j_n(\beta)h_n^{(1)}(\xi) \tag{63}$$

$$b_{mn} = \frac{2n+1}{n(n+1)}\frac{(n-m)!}{(n+m)!}[P_n^m(0)]^2 \tag{64}$$

$$b_n' = [\beta j_n(\beta)]'[\xi h_n^{(1)}(\xi)]' \tag{65}$$

$$\beta = 2\pi b/\lambda, \qquad \xi = 2\pi(b+a)/\lambda = (b+a)\beta/b. \tag{66}$$

The remaining properties may be similarly reduced. The solution given here may be compared to those obtained by Hallén[19] and Storer.[20]

B. Electrically Small Cases

1) *The Electrically Small Loaded Antenna:* If $\beta_1 \ll 1$, $\xi_1 \ll 1$, and $\xi_2 \ll 1$, then all terms in the current distribution and the input impedance formulas become negligible compared to the $m=0$ term. Thus

$$I(\phi') \simeq I(0) = V^i/Z^i \tag{67}$$

and

$$Z^i = 2880\pi^6\left(\frac{K_m}{K_m+2}\right)^2(b/\lambda)^4 - i240\pi^3K_m\frac{b}{\lambda}\sum_{n=1}^{\infty}\frac{2n+1}{nK_m+n+1}\frac{n+1}{n} \cdot [P_{n+1}(0)]^2(\beta_1/\xi_1)^n. \tag{68}$$

2) *The Electrically Small Antenna in Free Space:* If $K_m = K_e = 1$, in (68), then the case of an electrically small antenna in free space is obtained. This gives for the input impedance

$$Z^i = 320\pi^6\left(\frac{b}{\lambda}\right)^4 - i240\pi^2\frac{b}{\lambda}\left[\ln\frac{8b}{a} - 2\right]. \tag{69}$$

C. Results

It is to be noticed for the electrically small case that as $K_m \to \infty$ the radiation resistance as given by the real part of (68) tends to 9 times that of an antenna of the same size in free space.[6] However, since the antenna is required to be electrically small, then the ratio b/λ tends to zero. In such a case the absolute value of the input impedance also tends to zero.

Fig. 3 illustrates the behavior of the input impedance for loading material assumed lossless. The curves for the general size antenna indicate regions of maxima or resonances. These curves also contrast markedly to the curves obtained by assuming that the antenna is electrically small.

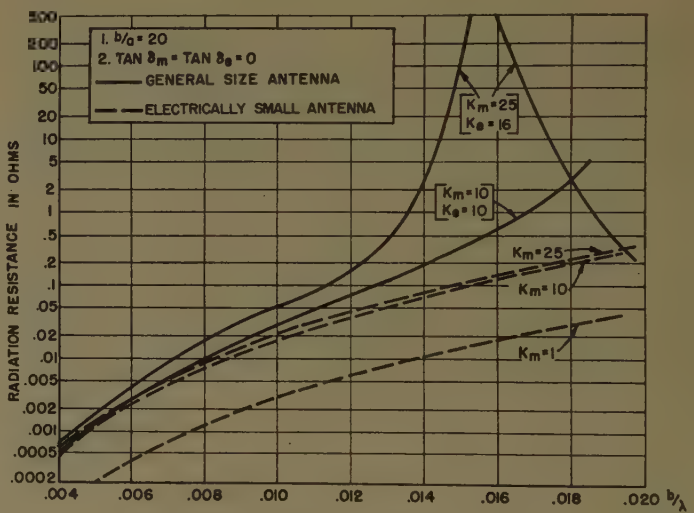


Fig. 3—Radiation resistance vs b/λ.

X. Acknowledgment

The author is indebted to Dr. P. R. Karr and E. E. Metzger for helpful discussions concerning various phases of the antenna problem, and wishes to express his appreciation for the computational aid rendered by Miss Isabelle Arsham.

[19] E. Hallén, "Theoretical investigations into the transmitting and receiving qualities of antennas," *Nova Acta Regiae Soc. Sci. Upsaliensis*, ser. 4, vol. 11, pp. 1–44; November, 1938.

[20] J. E. Storer, "Impedance of thin-wire loop antennas," *Commun. and Electronics*, no. 27, pp. 606–619; November, 1956.

The Conductance of Dipoles of Arbitrary Size and Shape*

K. FRÄNZ† AND P. A. MANN†

Summary—**The real part of either the impedance or the admittance of dipoles of arbitrary size and shape can be computed rigorously without solving a boundary value problem of a partial differential equation. In analogy to a well-known method of potential theory, fields of standing waves can be generated by integrals over current filaments so that for a given frequency there exist dipole shaped surfaces normal to the electric field surrounded by distant surfaces of vanishing electric field strength. Boundaries of perfect conductors may be supposed to coincide with a dipole shaped surface and a distant closed surface. The transients of such fields of standing waves are intimately related to the steady state of the free radiating dipole, since, before the first waves reflected from the distant enclosure have come back, the dipole cannot know whether or not it is enclosed. Corresponding to the type of current filament, either the resistance, or the conductance, of the radiating dipole can be calculated by direct integrations, while the shape of the dipole is determined by an ordinary differential equation of first order. As an example, we compute a family of dipoles that all have the same conductance $G=(254\ \Omega)^{-1}$ and a length $2h$ between limits $\lambda/2 \leq 2h \leq 1.36\cdot\lambda/2$.**

I. Introduction

THE electromagnetic field of a perfectly-conducting symmetric-transmitting dipole satisfies Maxwell's equations, a source condition at the input region, a boundary condition on the dipole surface and a radiation condition at infinity. The familiar antenna theories of Schelkunoff and Hallén reduce the problem of calculating dipole impedances to solving either an infinite system of linear equations, or an integral equation which can be solved by approximation methods. Both these theories calculate the impedance of a given antenna, for instance a biconical or a cylindrical one, as a function of frequency.[1,2]

It can be shown that for a given frequency the real part of either the impedance, or the admittance, of dipoles of finite length and thickness can be calculated by means of direct integrations, the dipole shape being determined by an ordinary differential equation of first order. This implies only numerical computations that can be performed on a digital machine, with a modest effort, to any desired degree of approximation.

Our method is similar to a well-known method of potential theory: Electrodes may be adapted to potential fields given by integrals over arbitrary charge distributions. With trivial modifications this same method can be applied to the fields inside cavities, *i.e.*, fields of standing waves, but not to complex radiation fields. This is due to the fact that, in a field of standing waves, a surface which is normal to the electric field at a given instant remains so at all times, while in a radiation field near an antenna, the direction of the electric vector varies during a cycle. Fortunately, the steady state of the radiation field of the dipole is intimately related to the transients of the fields of standing waves which are generated if this dipole is surrounded by a distant perfectly reflecting conductor. This is evident since, before the first waves reflected from the distant enclosure have come back, the dipole cannot know whether or not it is enclosed. By choosing a sufficiently distant enclosure, the delay between the beginning of the excitation of the enclosed dipole and the arrival of the first reflected waves can be made arbitrarily large. Because of radiation damping, it takes only a finite time to build up a radiation field and we obtain not only transients of the radiation field but, in the limit, its steady state.[3,4]

In Section II we establish (2), an analytical relation between the susceptance of the enclosed dipole and the admittance of the free radiating dipole. The susceptance does not appear explicitly in this equation but is first represented by a series of partial fractions. The admittance of the radiating dipole is then expressed in terms of the residues of this series and of the differences between resonance frequencies of the enclosed dipole. They can be calculated by direct integrations if the fields of standing waves are generated by current filaments according to (3). These results are established in Sections IV and V. The shape of the dipoles is given by the ordinary differential equation (19) which is integrated in Section VI for the case of sinusoidal current filaments of length $2l=\lambda/2$. Some dipole contours of this type are reproduced in Fig. 1.[3,4]

II. A Relation Between the Steady State of Radiation Fields and Transients of Fields of Standing Waves

The admittance Y_c of the enclosed dipole is a pure susceptance that can be written as a series of partial fractions

$$Y_c(p) = \sum_{-\infty}^{\infty} \frac{b_n}{p - j\omega_n} \qquad (1)$$

* Manuscript received by the PGAP, October 13, 1958; revised manuscript received, July 27, 1959.

† Telefunken Res. Inst., Ulm, Germany.

[1] S. Schelkunoff, "Advanced Antenna Theory," John Wiley and Sons, Inc., New York, N. Y.; 1952.

[2] R. W. P. King, "The Theory of Linear Antennas," Harvard University Press, Cambridge, Mass.; 1956.

[3] K. Fränz, "Bemerkungen zur Theorie des Strahlungsfeldes und der Impedanz von Antennen," *Arch. elect. Übertragung.*, vol. 10, pp. 269–273; July, 1956.

[4] K. Fränz, P. A. Mann, and J. Vocolides, "Der Wirkleitwert von Dipolen endlicher Länge und Dicke," *Arch. elect. Übertragung.*, vol. 12, pp. 49–53; February, 1958.

Also see K. Fränz and E. Henze, "Der Wirkwiderstand von Dipolen endlicher Länge und Dicke," *Arch. elect. Übertragung.*, vol. 13; October, 1959 (in print).

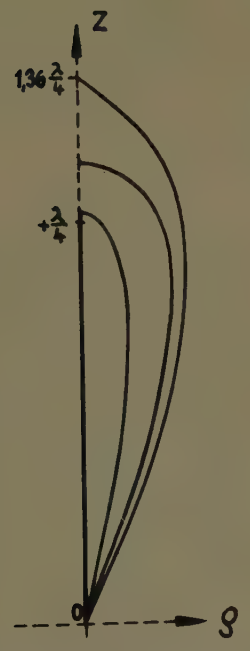


Fig. 1—Meridional sections of dipole arms. $G=254\ \Omega^{-1}$.

with real values $b_n=b_{-n}>0$, $\omega_n=-\omega_{-n}$ and $n\neq 0$.[5]

Let $u(t)$ be the voltage applied to the terminals of the dipole with

$$u(t)=\begin{cases}e^{p_0 t}\\0\end{cases}\qquad \text{for } t\begin{cases}>0\\<0;\end{cases}$$

it can alternatively be represented by the Fourier integral

$$u(t)=\frac{1}{2\pi j}\int_{-j\infty+\gamma}^{j\infty+\gamma}\frac{e^{pt}}{p-p_0}\,dp;\qquad \gamma>0.$$

The current $i(t)$ at the terminals of the enclosed dipole excited by the voltage $u(t)$ is

$$i(t)=\frac{1}{2\pi j}\int_{-j\infty+\gamma}^{j\infty+\gamma}Y_c(p)\,\frac{e^{pt}}{p-p_0}\,dp$$
$$=\sum_{-\infty}^{\infty}\frac{b_n(e^{j\omega_n t}-e^{p_0 t})}{j\omega_n-p_0}.$$

Let the dipole be confined to the vicinity of the origin, and let r_c be the minimum distance of the enclosure from the origin. If we first let r_c go towards infinity and then t,

$$i(t)e^{-p_0 t}\to Y(p_0)\qquad (2a)$$

approaches the admittance $Y(p_0)$ of the free radiating dipole.

With $r_c\to\infty$, the difference between resonance frequencies $\Delta_\omega=\omega_{n+1}-\omega_n$ approaches zero and so do the residues b_n. For imaginary frequencies $p_0=j\omega$, we obtain[6] (see Fränz, *et al.*[3,4]).

$$Y(j\omega)=\lim_{t=\infty}\left[\lim_{r_c=\infty}i(t)e^{-j\omega t}\right]$$
$$=\pi\lim_{r_c=\infty}\frac{b_n}{\Delta_\omega}+j2\omega\int_0^\infty\frac{\lim\limits_{r_c=\infty}\dfrac{b_\alpha}{\Delta_\alpha}}{\alpha^2-\omega^2}\,d\alpha\qquad (2b)$$

$\lim b_n/\Delta_\omega$ is an even analytical function of frequency.

The integral is understood to be Cauchy's principal value. As an example, we might substitute for Y_c the susceptance of an open ended cable; we would then correctly find, as radiation admittance of a cable of infinite length, the inverse of the characteristic impedance of this cable.

An analogous formula can be derived for the impedance of an antenna.[3,4]

III. Standing Waves Generated by Integrals over Current Distributions

The next step consists of generating fields of standing waves by means of integrals over arbitrary current distributions. We define a vector-potential A_z in cylindrical coordinates ρ, z, ϕ through

$$A_z=\frac{1}{4\pi}\int_{-l}^{l}\left[i_1(\zeta)\frac{\cos kr}{r}+i_2(\zeta)\frac{\sin kr}{r}\right]d\zeta$$
$$ck=\omega,\qquad (3)$$

(compare to Fig. 2) and the corresponding electromagnetic fields, as usual through equations

$$H_\phi=-\frac{\partial A_z}{\partial\rho}\qquad (4)$$

$$j\omega\epsilon_0\rho E_z=\frac{\partial(\rho H_\phi)}{\partial\rho}\qquad (5a)$$

$$j\omega\epsilon_0\rho E_\rho=-\frac{\partial(\rho H_\phi)}{\partial z}.\qquad (5b)$$

In order to generate standing waves, i_1 and i_2 must be chosen to be real; with $i_2=-ji_1$ the kernel under the integral would turn out to be e^{-jkr}/r and the result would be a radiation field satisfying a radiation condition at infinity.

In case we are able to substitute currents in (3) such that there exist dipole shaped surfaces normal to the electric field with biconical input regions surrounded by distant closed surfaces of vanishing electric field-strength, and in case we are able to evaluate lim b_n/Δ_ω in (2b), we can, indeed, compute the real part of the admittance of a dipole of finite length and thickness at the frequency $\omega=kc$ (c = velocity of light).

In the present paper, we shall compute a family of dipoles that all have the same conductance $G=(254\Omega)^{-1}$. Their length, $2h$, is restricted to values $\lambda/2\leq 2h\leq 1.36\ \lambda/2$. (See Fig. 1.)

[5] K. Fränz, "Das Reaktanztheorem für beliebige Hohlräume," *Elec. Nachrichtentech.*, vol. 21, pp. 8–12; January, 1944.

[6] A proof of (2b) is given in Appendix I.

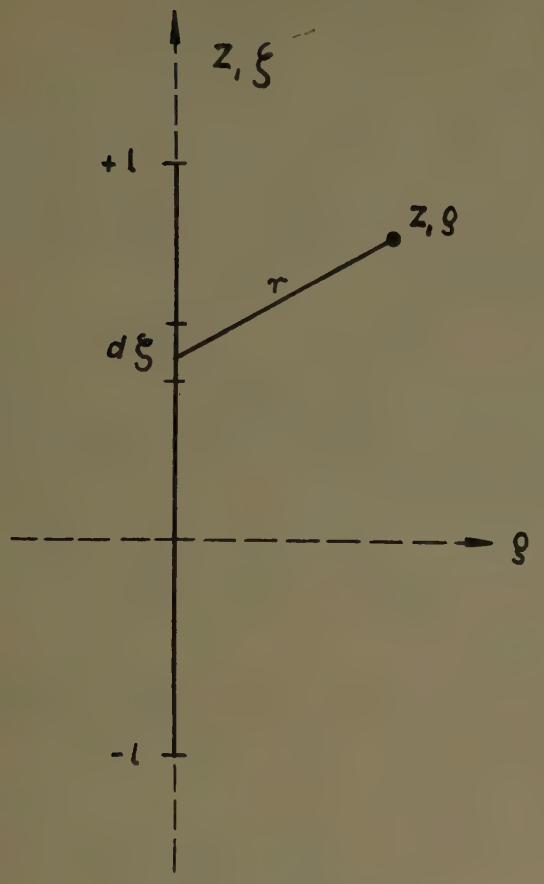


Fig. 2—Coordinates for (3).

IV. The Asymptotic Behavior of Resonance Frequencies of Large Cavities

We do not intend to present our theory in as general a form as possible, but, rather, to demonstrate that precise numerical results can be obtained with a modest effort. This is particularly easy if, in (3), we choose $i_2 = \alpha i_1$, α being a real constant. In this case, the field at large distances from the origin ($kr \gg 1$) is of the type

$$E_\vartheta = -jZ_0 f(\theta)\left[\frac{\cos kr_0}{r_0} + \alpha\frac{\sin kr_0}{r_0}\right] \tag{6a}$$

$$H_\phi = f(\theta)\left[\frac{\sin kr_0}{r_0} - \alpha\frac{\cos kr_0}{r_0}\right] \tag{6b}$$

$$Z_0 = \sqrt{\mu_0/\epsilon_0}.$$

According to (6a), the electric field vanishes on equidistant spheres, their mutual distance being equal to $\lambda/2$. Conducting reflectors must coincide with these spheres. An increase $\delta k = \pi/r_c$ in (3) causes a shift from one sphere to the next, while the variation of the field in the vicinity of the dipole can be made infinitesimal with δk by making r_c infinite. Let $\omega_n = k_n c$ be a resonance frequency, then the next resonance frequency is

$$\omega_{n+1} = \omega_n + \Delta_\omega = k_n c + \pi c/r_c. \tag{7}$$

With increasing distance r_c, the difference Δ_ω between resonance frequencies becomes arbitrarily small.

It is well known that the asymptotic behavior of the eigenfrequencies of arbitrary cavities is very simple, the number of eigenfrequencies below a given sufficiently-large frequency depending only on the volume inside the cavity, not on its shape. Since our dipole surfaces are surfaces of revolution, for reasons of symmetry only a fraction of the eigenfrequencies of the cavity are excited and appear as resonance frequencies in (1).

If the quotient between the two currents i_1 and i_2 is not independent of ζ, there still exist distant nearly spherical surfaces of vanishing electric field strength whose mutual distance is $\lambda/2$. Eq. (7) is equivalent to

$$\frac{\delta\omega}{\omega_n} = \frac{\lambda_n}{2r_c}. \tag{8}$$

V. The Asymptotic Behavior of Residues for Large Cavities

If, in (3), we substitute currents i_1 that have a node at the input region, $\omega = kc$ is a zero of the susceptance and a pole of the reactance. If we substitute currents i_1 that have continuous first derivative $\delta i_1/\delta\zeta$ at the input region, then $\omega = kc$ is a zero of the reactance and a pole of the susceptance. Particularly simple formulas result for $i_1 = i_0 \cos k\zeta$, $l = \lambda/4$, $i_2 = \alpha i_1$, α being a real constant.

The reactance is of the form

$$jx_c(j\omega) = j(\omega - \omega_n)\left(\frac{\partial x_c}{\partial\omega}\right)_{\omega=\omega_n} + \cdots;$$

thus, the residue b_n of the susceptance at $\omega = \omega_n$ is

$$b_n = \left(\frac{\partial x_c}{\partial\omega}\right)^{-1}_{\omega=\omega_n}. \tag{9}$$

It can be shown that the variation $\delta x_c/\delta\omega$ of a cavity reactance with frequency can be expressed through its variation for infinitesimal deformations of the cavity. In the case of a field given by (3), the reactance is a function of the product ωl; we therefore have

$$\omega\frac{\partial x_c}{\partial\omega} = l\frac{\partial x_c}{\partial l}. \tag{10}$$

The reactive power P inside the cavity is equal to

$$P = \omega\int(\mu_0 HH^* - \epsilon_0 EE^*)dv. \tag{11}$$

In theoretical physics the integral itself is called the Lagrangian of the field. This formula is valid for both cavities and radiation fields. In a radiation field, the magnetic energy $\int\mu_0 HH^* dv$ and the electric energy $\int\epsilon_0 EE^* dv$ are divergent quantities if taken separately; but the Lagrangian is finite since for the field of a transmitting antenna, $\mu_0 HH^* - \epsilon_0 EE^*$ is asymptotically proportional to the inverse fourth power of the distance from the source.

Keeping the current i_0 in the terminals constant, the variation of the reactive power also gives the variation of the reactance.

$$\delta P = i_0^2 \delta x_c. \tag{12}$$

The variation δP of the reactive power for infinitesimal deformations can easily be computed (see Fränz, *et al.*[3,4])

$$\delta P = \omega \int (\mu_0 HH^* - \epsilon_0 EE^*) \delta n d\sigma. \tag{13}$$

The integral must be extended over the whole surface of the cavity, *i.e.*, in our case over the distant sphere and the dipole itself. In this formula δn is the normal displacement of the surface element $d\sigma$ due to the infinitesimal deformation of the cavity. Eq. (13) is valid if the walls of the cavity are perfectly *conducting* reflectors and if not only the variation of the volume remains infinitesimal, but all displacements δn. It is not valid for displacements of the reflecting open end of a cable or for introducing into cavities wires of infinitesimal diameter, but finite length.

$\mu_0 HH^*$ is equal to the pressure of the magnetic field on the surface element $d\sigma$ in the direction of the outer normal, $\epsilon_0 EE^*$ represents the drag of the electric field on $d\sigma$ directed towards the interior of the cavity.

In case of an infinitesimal variation δl, the displacement of the dipole surface can be made arbitrarily small compared to the displacements δn of the distant enclosure, their ratio being of the order l/r_c. Since the electric field vanishes on the distant sphere, that which contributes to variations of P with δl is only

$$\delta P = \omega \delta n \int \mu_0 HH^* d\sigma = r_c \delta\omega \int \mu_0 HH^* d\sigma \tag{14}$$

taken over the sphere. Combining

$$G = \pi \lim_{r_c = \infty} \frac{b_n}{\Delta_\omega} \tag{15}$$

with (9), (10), (12) and (14), we obtain

$$G^{-1} = Z_0 i_0^{-2} \lim_{r_c = \infty} \int HH^* d\sigma; \qquad Z_0 = \sqrt{\mu_0/\epsilon_0}. \tag{16}$$

Thus the conductance G of the radiating antenna is essentially determined by the total radiation pressure on the distant enclosure of the cavity.

VI. The Differential Equation of the Dipole Contour and its Solution for a Sinusoidal Current Filament of Length $2l = \lambda/2$

The differential equation of the electric lines of force is

$$E_\rho dz - E_z d\rho = 0. \tag{17}$$

Comparing with (5a) and (5b), we find its solution

$$\rho H_\phi = \text{const.} \tag{18}$$

The dipole contour is orthogonal to the lines of force and, therefore satisfies equation

$$E_\rho d\rho + E_z dz = 0. \tag{19}$$

We have to integrate (19) for currents

$$i_1(\zeta) = i_0 \cos \pi\zeta/2l, \quad 2l = \lambda/2, \quad i_2 = \alpha i_1. \tag{20}$$

The constant α turns out to be determined by the condition that the input region must be biconical. We shall see that this happens only for $\alpha = -\pi/2$; this may be established by analysis of the lines of force. With currents (20) we find (see Schelkunoff[7]), as equation of the lines of force,

$$4\pi\rho H_\phi/i_0 = \sin kr_1 + \sin kr_2 - \alpha(\cos kr_1 + \cos kr_2). \tag{21}$$

In the vicinity of the input region $r_0 \sim 0$, the lines of force must be concentric circles with center $r_0 = 0$ if there exists a family of dipole-shaped contours with biconical input regions. After introducing the deviations of kr_1 and kr_2 from $\pi/2$,

$$kr_1 = \frac{\pi}{2} - \Delta_1, \qquad kr_2 = \frac{\pi}{2} - \Delta_2,$$

we obtain

$$4\pi\rho H_\phi/i_0 = \cos\Delta_1 + \cos\Delta_2 - \alpha(\sin\Delta_1 + \sin\Delta_2)$$
$$\approx 2 - \frac{\Delta_1^2 + \Delta_2^2}{2} - \alpha(\Delta_1 + \Delta_2) + \cdots.$$

From the geometry of Fig. 3, we see that

$$(kr_1)^2 = \frac{\pi^2}{4} + (kr_0)^2 + \pi kr_0 \cos\theta = \left(\frac{\pi}{2} - \Delta_1\right)^2$$

$$(kr_2)^2 = \frac{\pi^2}{4} + (kr_0)^2 - \pi kr_0 \cos\theta = \left(\frac{\pi}{2} - \Delta_2\right)^2,$$

or

$$2(kr_0)^2 = \Delta_1^2 + \Delta_2^2 - \pi(\Delta_1 + \Delta_2)$$

and

$$4\pi\rho H_\phi/i_0 = 2 - (\Delta_1^2 + \Delta_2^2)\left(\frac{1}{2} + \frac{\alpha}{\pi}\right) + \frac{2\alpha}{\pi}(kr_0)^2 + \cdots.$$

Evidently $\alpha = -\pi/2$ is sufficient to make $4\pi\rho H_\phi/i_0$ a function of r_0 alone and independent of θ. Since $\Delta_1^2 + \Delta_2^2$ is not independent of θ, this condition is also necessary.

Before integrating (19) for the dipole contour, we compute the conductance G of our family of dipoles. Substituting $i(\zeta) = i_0 \cos \pi\zeta/2l$ and $\alpha = -j$, we obtain a radiation field of the form

$$H_\phi^{(1)} = i_0 f(\theta) \frac{\sin kr_0 + j \cos kr_0}{r_0}$$

instead of our field of standing waves

$$H_\phi^{(2)} = i_0 f(\theta) \frac{\sin kr_0 - \alpha \cos kr_0}{r_0}$$

[7] S. Schelkunoff, "Electromagnetic Waves," D. Van Nostrand Co., Inc., New York, N. Y., p. 371; 1943.

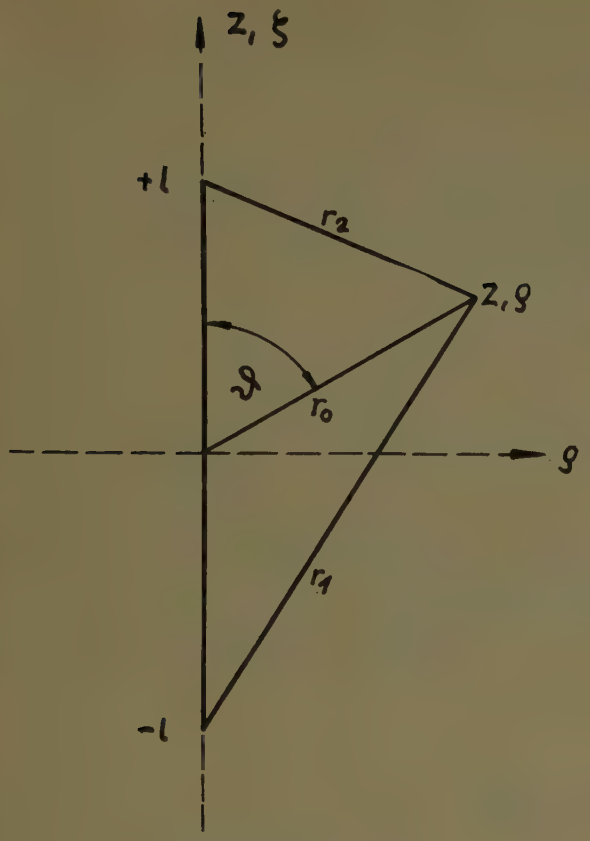


Fig. 3—Coordinates for (21).

with $\alpha = -\pi/2$. From the elementary theory of radiation resistance we know

$$\sqrt{\frac{\mu_0}{\epsilon_0}} \int |H^{(1)}|^2 d\sigma = i_0^2 R_0 \quad \text{with} \quad R_0 = 73.2\ \Omega. \tag{22}$$

The similar integral of (16) must be taken for a sphere of vanishing electric field $E_\vartheta = 0$; *i.e.*, for a sphere where $\sin kr_0 - \alpha \cos kr_0$ takes its maximum $\sqrt{1+\alpha^2}$, or

$$i_0^2 \frac{\partial x_0}{\partial \omega} = r_0 \int \mu_0 |H^{(2)}|^2 d\sigma = i_0^2 R_0 \frac{r_0}{c} (1 + \alpha^2).$$

For the conductance G, we obtain

$$G^{-1} = R_0 \left(1 + \frac{\pi^2}{4}\right) = 254\ \Omega.$$

In order to integrate differential equation (19) for the dipole contours, we first transform (18) for the lines of force.

$$\sin kr_1 + \sin kr_2 + \frac{\pi}{2}(\cos kr_1 + \cos kr_2) = \text{const}$$

is equivalent to

$$\cos\left(k\frac{r_1 + r_2}{2} - \psi\right) \cos k\left(\frac{r_1 - r_2}{2}\right) = \text{const} \tag{23}$$

with $\psi = \operatorname{arctg} 2/\pi$.

This latter form (23), of (21), suggests the use of confocal ellipses $r_1 + r_2 = \text{const}$ and hyperbolas $r_1 - r_2 = \text{const}$ as coordinates. By conformal mapping of the z, ρ-plane on a ξ, η-plane, the images of the lines of force and the dipole contours remain orthogonal. A suitable transformation is

$$z + j\rho = \frac{\lambda}{4} \cosh(\xi + j\eta)$$

$$k\frac{r_1 + r_2}{2} = \frac{\pi}{2} \cosh \xi; \qquad k\frac{r_1 - r_2}{2} = \frac{\pi}{2} \cos \eta.$$

The images of the lines of force are the curves

$$F(\xi, \eta) = \cos\left(\frac{\pi}{2} \cosh \xi - \psi\right) \cos\left(\frac{\pi}{2} \cos \eta\right) = \text{const}.$$

The images of the dipole contours are the corresponding orthogonal trajectories satisfying equation

$$\frac{\partial F}{\partial \eta} d\xi - \frac{\partial F}{\partial \xi} d\eta = 0. \tag{24}$$

The substitution $\cosh \xi = u$, $\cos \eta = v$ gives

$$\frac{\cot\left(\frac{\pi}{2} u - \psi\right)}{u^2 - 1} du + \frac{\cot\left(\frac{\pi}{2} v\right)}{v^2 - 1} dv = 0.$$

$$1 \le u < \infty; \qquad -1 \le v \le 1.$$

This equation is integrated by

$$\int_u^{u(1)} \frac{\cot\left(\frac{\pi}{2} u - \psi\right)}{u^2 - 1} du = \int_v^1 \frac{\cot \frac{\pi}{2} v}{1 - v^2} dv. \tag{25}$$

We have fixed the constant of integration so that $u = u(1)$ for $v = 1$. The integral over v has a logarithmic singularity for $v = 0$. For

$$1 < u \le u(1) \le 1 + \frac{2}{\pi}\psi = 1.36$$

we obtain

$$\cot\left(\frac{\pi}{2} u - \psi\right) \ge 0, \qquad \frac{du}{dv} > 0$$

and v decreases simultaneously with u. So v must approach 1 to make the integral over v infinite. The point $u = 1$, $v = 0$ corresponds to the origin of the z, ρ-plane. The distance $2h$ between the extremes of the two dipole arms is given by $2h = \lambda u(1)/2$. All the orthogonal contours satisfying the following inequality

$$1 < 4h/\lambda = u(1) < 1 + \frac{2}{\pi}\psi = 1.36$$

terminate conically at the origin of the z, ρ-plane and so correspond to dipole-shaped surfaces of revolution with biconical input regions. Since u decreases monotonously with v, all dipole contours remain inside that ellipse $r_1 + r_2 = \text{const}$ which is tangent to the contour at its extremes.

The calculus of dipole resistance instead of conductance is quite similar to the proceedings of this paper.

Appendix I

Proof of (2b)

In order to prove (2b), we must evaluate the sum

$$i(t)e^{-p_0 t} = \sum_{-\infty}^{\infty} \frac{b_n}{\Delta_\omega} \frac{e^{(j\omega_n - p_0)t} - 1}{j\omega_n - p_0} \Delta_\omega$$

for infinite distance of the enclosure of the dipole. This sum approaches an integral, since b_n and the difference between resonance frequencies $\Delta_\omega = \omega_{n+1} - \omega_n$ both approach zero, $\lim b_n/\Delta_\omega$ being an even analytical function of frequency. After separation of real and imaginary parts, we obtain

$$Y(j\omega) = \lim_{t=\infty} \left[\int_{-\infty}^{\infty} \lim_{r_c=\infty} \frac{b_n}{\Delta_\omega} \frac{\sin(\alpha-\omega)t}{\alpha-\omega} d\alpha + j \int_{-\infty}^{\infty} \lim_{r_c=\infty} \frac{b_n}{\Delta_\omega} \frac{1-\cos(\alpha-\omega)t}{\alpha-\omega} d\alpha \right].$$

The limit of the integral for the real part can easily be calculated for infinite t according to formulas well-known from the theory of Fourier integrals.[8] In order to evaluate the integral for the imaginary part, we observe that with the exception of an infinitesimal interval near $\omega = \omega_0$,

$$\int_{-\infty}^{\infty} \lim \frac{b_n}{\Delta_\omega} \frac{1-\cos(\alpha-\omega)t}{\alpha-\omega} d\alpha = 2\int \lim \frac{b_n}{\Delta_\omega} \frac{\sin^2 \frac{\alpha-\omega}{2} t}{\alpha-\omega} d\alpha = 2\int \lim \frac{b_n}{\Delta_\omega} \frac{\cos^2 \frac{\alpha-\omega}{2} t}{\alpha-\omega} d\alpha = \int_{-\infty}^{\infty} \lim \frac{b_n}{\Delta_\omega} \frac{d\alpha}{\alpha-\omega}.$$

This gives (2) since the real part of Y; *i.e.*, $\lim b_n/\Delta_\omega$, is an even function of ω.

$$Y(j\omega) = \pi \lim_{r_c=\infty} \frac{b_n}{\Delta_\omega} + j2\omega \int_0^{\infty} \lim_{r_c=\infty} \frac{b_n}{\Delta_\omega} \frac{d\alpha}{\alpha^2 - \omega^2}.$$

Our evaluation of the imaginary part may be replaced by the simple remark that our result represents only a well-known relation between the real and imaginary parts of any positive function.[9]

Appendix II

Proof of (13)

The reactive power P contained in a cavity is given by (11). Let ω be a constant frequency and δn an infinitesimal displacement of the element $d\sigma$ normal to the surface of the cavity. We suppose that the input region and the current i_0 in the terminals are not varied. In our case of deformations caused by variation δl of a scale factor, a biconical input region located at the origin remains, indeed, biconical, and the angle of the cones remains fixed. The variation δP of the reactive power is equal to

$$\delta P = \omega \int_{II} (\mu_0 H H^* - \epsilon_0 E E^*) \delta_n d\sigma + \omega \int_I [(\mu_0 H \delta H^* - \epsilon_0 E \delta E^*) + (\mu_0 H^* \delta H - \epsilon_0 E^* \delta E)] dv.$$

Integral II is to be taken over the complete surface of the cavity, integral I over the original volume; δE and δH are the variations of the electromagnetic field. We now proceed to show that integral I is of a minor order to that of integral II. The fields satisfy Maxwell's equation

$$j\omega\epsilon_0 E = \text{rot } H.$$

We therefore have

$$k^2(\mu_0 H \delta H^* - \epsilon_0 E \delta E^*) = \mu_0(k^2 H \delta H^* + \text{rot } H \text{ rot } \delta H^*).$$

Because of

$$\text{div}(\text{rot } H \times \delta H^*) = \delta H^* \text{ rot rot } H - \text{rot } H \text{ rot } \delta H^*$$

and

$$\text{rot rot } H = -\Delta H$$

we find, for integral I,

$$\frac{\mu_0}{k^2} \int [\delta H^*(k^2 H + \Delta H) + \delta H(k^2 H^* + \Delta H^*) dv + k^2 \int (\text{rot } H \times \delta H^* + \text{rot } H^* \times \delta H)_n d\sigma.$$

The latter volume integral vanishes for all fields that satisfy Maxwell's equations. The surface integral is of a minor order to that of II since δH vanishes with δn and, simultaneously, rot $H \approx E$ becomes normal to $d\sigma$. Thus, for infinitesimal deformations, we may neglect I compared to II. We may check (13), substituting variations of the length of a short circuited concentric cable. The analogy of (13) with a formula for the variation of the resonance frequency of a cavity with deformations of its walls, published in 1939 by J. Müller, is evident.[10]

Acknowledgment

The authors wish to thank E. Henze for his reading of this manuscript and for his programming of the integrals (25) for computation on a digital machine.

[8] E. T. Whittaker, and G. N. Watson, "A Course of Modern Analysis," Cambridge University Press, Cambridge, England, p. 188; 1952.

[9] H. W. Bode, "Network Analysis and Feedback Amplifier Design," D. Van Nostrand Co. Inc., New York, N. Y., p. 335; 1953.

[10] J. Müller, "Untersuchungen über elektromagnetische Hohlräume," *Z. Hochfrequenz.*, vol. 54, pp. 157–161; 1939.

The Launching of Surface Waves by a Parallel Plate Waveguide*

C. M. ANGULO† AND W. S. C. CHANG‡

Summary—**The excitation of the lowest TM surface wave in grounded dielectric slab by a terminated parallel plate waveguide is discussed. The ground plane is the continuation of the lower plate of the waveguide and the infinite dielectric slab is partially filling the waveguide. The thickness of the slab, the height of the parallel plate waveguide, and the frequency are such that only the lowest slow wave can propagate in the partially filled waveguide and the grounded dielectric slab.**

The Fourier transform of the field scattered by the termination of the upper plate of the waveguide is found by means of the Wiener-Hopf technique and the far fields obtained by the method of steepest descents. The percentage of power reflected back into the waveguide, of power transmitted to the surface wave in the slab, and of power radiated into the open space are plotted vs the thickness of the slab for different heights of the waveguide and $\epsilon=2.49$.

This method of excitation is found to be very efficient. If the dimensions of the waveguide and the slab remain within a considerably wide range, the efficiency obtained for a given frequency is very close to the optimum. Therefore, the adjustments for maximum efficiency are not critical.

Introduction

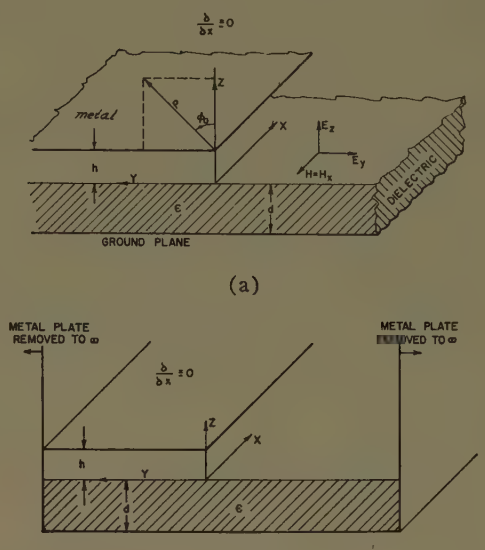
THERE has been great interest in surface wave structures in recent years as open waveguides and as antennas. Therefore, numerous papers have been published on the propagation in such structures as well as on means of excitation. The reader is referred to Zucker's paper[1] which contains a very complete bibliography on the subject as well as an excellent survey.

A very convenient practical way of launching a surface wave into a dielectric slab is a parallel plate waveguide [see Fig. 1(a)]. The input energy is contained in the dominant TM ($H_y=0$) mode of the partially filled waveguide propagating from $y=+\infty$ to $y=0$. The dimensions of the guide and the thickness of the slab are restricted to the range for which only one surface wave (the lowest symmetric mode)[2] exists along the slab and only one mode (the dominant TM) can propagate inside the partially filled waveguide. These conditions are:[3]

$$Kd < \pi(\epsilon)^{-1/2} \tag{1a}$$

$$Kh < \arctan\left\{-(\epsilon)^{-1/2}\tan\left[(\epsilon)^{1/2}Kd\right]\right\} \tag{1b}$$

$$0 < \arctan\left\{-(\epsilon)^{-1/2}\tan\left[(\epsilon)^{1/2}Kd\right]\right\} < \pi \tag{1c}$$



(a)

(b)

Fig. 1.

where $K=\omega(\mu_0\epsilon_0)^{1/2}$, ϵ is the relative permittivity, d is the thickness of the dielectric, and $d+h$ is the height of the parallel plate waveguide as indicated in Fig. 1(a).

Because of the discontinuity at $y=0$ where the upper plate is terminated, the energy incident upon the discontinuity will be partly reflected back into the waveguide, partly transmitted to the surface wave in the grounded dielectric slab and partly radiated. We are interested in finding the three power ratios for different values of Kd and Kh.

Since the structure shown in Fig. 1(a) does not vary in the x direction and the incident wave is the lowest TM mode in the partially filled waveguide, all the fields excited will be independent of x and will have $E_x=H_y=H_z=0$.

The structure shown in Fig. 1(a) is regarded mathematically as a homogeneous parallel plate air waveguide (with walls at $y=\pm\infty$) and extending from $z=0$ to $z=+\infty$ connected to a homogeneous parallel plate waveguide of length d, filled with dielectric of relative permittivity ϵ (also with walls at $y=\pm\infty$) and terminated by an electric wall at $z=-d$. Inside the first waveguide there is an obstacle, a semi-infinite perfectly conducting plane, placed at $z=h$ from $y=0$ to $y=+\infty$. Fig. 1(b) illustrates the above description. By removing the $y=$ const. walls to infinity, the structure of Fig. 1(a) is obtained.

The modal analysis of a parallel plate waveguide with walls at infinity represents the transversal fields (E_y and H_z, in our case) in terms of their Fourier transforms in the cross-section variable (y, in our case). In the successive sections we shall proceed as follows:

* Manuscript received by the PGAP, June, 20 1958; revised manuscript received July 20, 1959. The work described in this paper has been sponsored by the Air Force Cambridge Research Center under Contract AF 19(604)-1391 with Brown University.

† Brown University, Providence, R. I.

‡ Ohio State University, Columbus, Ohio.

[1] F. J. Zucker, "The guiding and radiation of surface wave," *Proc. Symp. on Modern Advances in Microwave Techniques*, Polytechnic Inst. of Brooklyn, Brooklyn, N. Y., pp. 403–435; November, 1954.

[2] C. M. Angulo and W. S. C. Chang, "On the Excitation of Surface Waves in Dielectric Slabs, Part I: Excitation by a Magnetic Line," Div. of Engrg., Brown Univ., Providence, R. I., Rept. No. AF 1391/2; June, 1956. AFCRC-TN-56-950, ASTIA Doc. No. AD 110141.

[3] C. M. Angulo and W. S. C. Chang, "The Launching of Surface Waves by a Parallel Plate Waveguide," Div. of Engrg., Brown Univ., Providence, R. I., Rept. No. 1391/5; April, 1957. AFCRC-TN-57-365, ASTIA Doc. No. AD 117059, pp. 38–46.

1) The equation relating the Fourier transforms of the fields at the semi-infinite obstacle will be derived by the modal analysis method.
2) The Wiener-Hopf technique will be applied to the solution of the equation obtained, and the exact fields will be expressed as the results of integrations in the complex plane.[4–6]
3) These integrals will be evaluated only at points far away from the discontinuity. The evaluation will be carried out by analyzing the relationship between the singularities of the integrands and the analytical forms of the far fields, which are known. In fact, for $y \ll 0$ the principal contribution on the surface of the slab must be the principal surface wave propagating along a grounded dielectric slab, and for $y \gg 0$ and $z < h$ the fields must be those of the dominant mode in a parallel plate waveguide partially filled with dielectric.

The Equation for the Fourier Transforms of the Fields at Plane $z = h$

We first separate H_z and E_y into the incident and scattered fields; in a second step, we find the expressions for the Fourier transforms of the scattered fields for $z < h$ and $z > h$; and finally we match the boundary conditions at $z = h$ in terms of the Fourier transforms.

Let us represent the fields everywhere as:

$$H_z = H_{0z} + \mathcal{H}_z \tag{2a}$$

$$E_y = E_{0y} + \mathcal{E}_y \tag{2b}$$

where E_{0y} and H_{0z} are the components of the dominant TM mode of the partially filled parallel plate waveguide and $\mathcal{H}_z$ and $\mathcal{E}_y$ are the scattered fields. For $-d < z < h$ we have

$$H_{0z} = \left\{ \frac{\cosh [K(z-h)s']}{\cosh (Khs')} u(z) + \frac{\cos [K(z+d)r']}{\cos (Kdr')} u(-z) \right\} \times \exp [jK(1+s'^2)^{1/2} y] \tag{3a}$$

$$E_{0y} = -j \frac{Ks' \tanh (Khs')}{\omega \epsilon \epsilon_0} \left\{ \frac{\sinh [K(z-h)s']}{\sinh (Khs')} u(z) - \frac{\sin [K(z+d)r']}{\sin (Kdr')} u(-z) \right\} \times \exp [jK(1+s'^2)^{1/2} y]. \tag{3b}$$

For $h < z$ we have

$$H_{0z} = 0 \tag{3c}$$

$$E_{0y} = 0. \tag{3d}$$

The time dependence is taken as $e^{j\omega t}$.

The function $u(z)$ represents Heaviside's unit step function, zero for negative argument and one for positive arguments. The quantities r' and s' are the modulus of the wave numbers in the OZ direction in the dielectric and in the air respectively normalized with respect to K for the incident mode. They are the solutions of the following equations

$$r'^2 + s'^2 = \epsilon - 1 \tag{4a}$$

$$\frac{\epsilon s'}{r'} = \frac{\tan (Kdr')}{\tanh (Khs')}. \tag{4b}$$

Figs. 2 and 3 are plots of r' and s' vs height of the waveguide for several thicknesses of the slab and $\epsilon = 2.49$.

We will find below the quantity s similar to s'. s represents the modulus of the wavenumber in the OZ direction, in the air, normalized with respect to K for the lowest TM surface wave in a grounded dielectric slab. A plot of s vs thickness of the slab has been published[7] for $\epsilon = 2.49$.

The scattered fields can be represented by their Fourier transforms:

$$\mathcal{H}_z(y, z) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} I(\eta, z) e^{-j\eta y} d\eta \tag{5a}$$

$$\mathcal{E}_y(y, z) = \frac{-1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} V(\eta, z) e^{-j\eta y} d\eta \tag{5b}$$

$$I(\eta, z) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} \mathcal{H}_z(y, z) e^{j\eta y} dy \tag{5c}$$

$$V(\eta, z) = \frac{-1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} \mathcal{E}_y(y, z) e^{j\eta y} dy. \tag{5d}$$

Maxwell's equations require that the transforms be solutions of the transmission line equations:

$$\frac{dV(\eta, z)}{dz} = -j\zeta Z I(\eta, z) \tag{6a}$$

$$\frac{dI(\eta, z)}{dz} = -j\zeta Y V(\eta, z) \tag{6b}$$

for $z > h$ and for $h > z > -d$.

In the air,

$$\zeta = \zeta_a = (K^2 - \eta^2)^{1/2} \tag{6c}$$

$$Y_a = \frac{1}{Z_a} = \omega \epsilon_0 (K^2 - \eta^2)^{-1/2}. \tag{6d}$$

In the dielectric,

$$\zeta = \zeta_d = (K^2 \epsilon - \eta^2)^{1/2} \tag{6e}$$

$$Y_d = \frac{1}{Z_d} = \omega \epsilon_0 \epsilon (K^2 \epsilon - \eta^2)^{-1/2}. \tag{6f}$$

[4] R. Paley and N. Wiener, "Fourier Transform in Complex Domain," Amer. Mathematical Soc. Colloquium Publications, vol. XIX, pp. 49–58; 1943.

[5] P. M. Morse and H. Feshbach, "Methods of Theoretical Physics," McGraw-Hill Book Co., Inc., New York, N. Y., vol. 1, pp. 960–992; 1953.

[6] For results of a similar analysis for the same configuration without the dielectric slab, see N. Marcuvitz, "Waveguide Handbook," McGraw-Hill Book Co. Inc., New York, N. Y., Radiation Laboratory Series, vol. 10, pp. 179–186; 1951.

[7] C. M. Angulo, "Diffraction of surface waves by a semi-infinite dielectric slab," IRE Trans. on Antennas and Propagation, vol. AP-5, pp. 100–109; January, 1957. See p. 102.

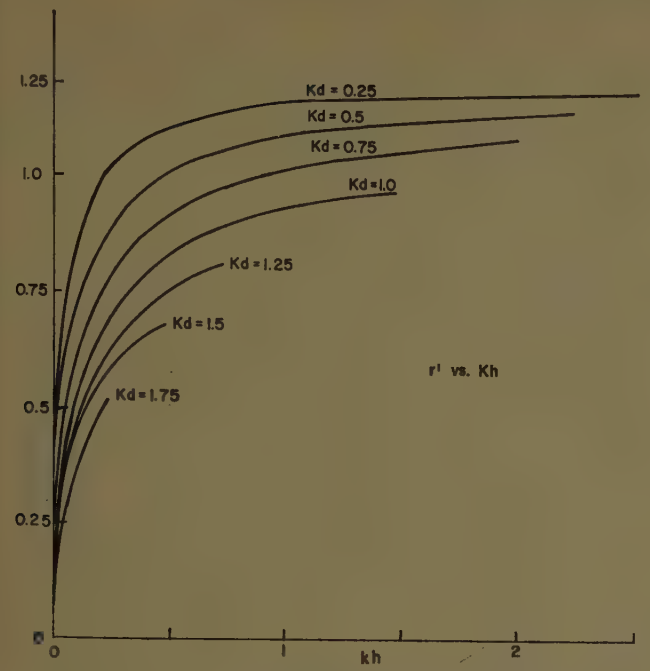


Fig. 2—r' vs Kh for $\epsilon=2.49$, for several thicknesses of the grounded slab.

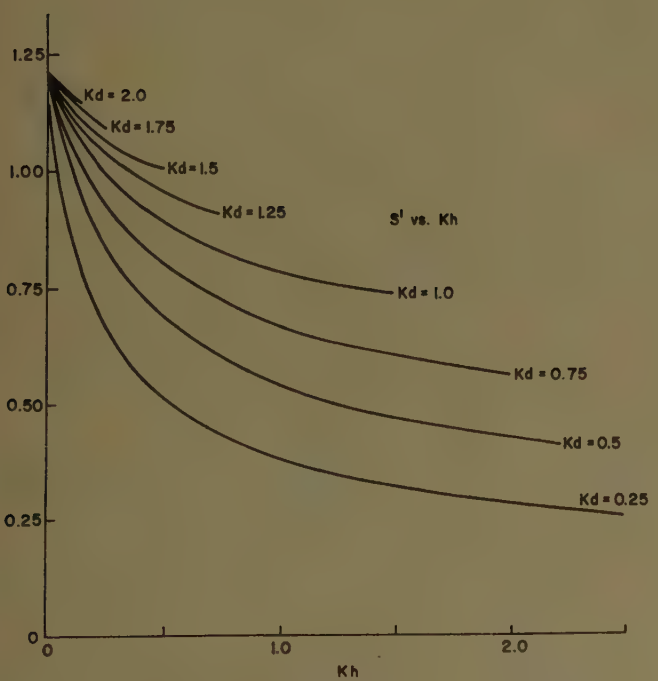


Fig. 3—s' vs Kh for $\epsilon=2.49$, for several thicknesses of the grounded slab.

If we recall Fig. 1, we see that the solutions for V and I can be written immediately from the theory of transmission lines for the two regions $z>h$ and $z<h$, as follows:

For $z>h$,

$$V(\eta, z) = V(\eta, h_+) \exp\{-j\zeta_a(z-h)\} \tag{7a}$$

$$I(\eta, z) = Y_a V(\eta, h_+) \exp\{-j\zeta_a(z-h)\}. \tag{7b}$$

For $h>z>0$,

$$V(\eta, z) = V(\eta, h_-)\cos\zeta_a(z-h) - jZ_aI(\eta, h_-)\sin\zeta_a(z-h) \tag{8a}$$

$$I(\eta, z) = I(\eta, h_-)\cos\zeta_a(z-h) - jY_aV(\eta, h_-)\sin\zeta_a(z-h). \tag{8b}$$

Finally, for $0>z>-d$,

$$V(\eta, z) = [V(\eta, h_-)\cos\zeta_a h + jZ_aI(\eta, h_-)\sin\zeta_a h] \cdot \frac{\sin\zeta_d(z+d)}{\sin\zeta_d d} \tag{8c}$$

$$I(\eta, z) = [I(\eta, h_-)\cos\zeta_a h + jY_aV(\eta, h_-)\sin\zeta_a h] \cdot \frac{\cos\zeta_d(z+d)}{\cos\zeta_d d}, \tag{8d}$$

where

$$I(\eta, h_-) = -jY_aV(\eta, h_-)\frac{\frac{\zeta_d}{\epsilon}\tan(\zeta_a h)\tan(\zeta_d d) - \zeta_a}{\frac{\zeta_d}{\epsilon}\tan(\zeta_d d) + \zeta_d\tan(\zeta_a h)}. \tag{8e}$$

The relationship between the values of V and I at $z=h_-$ and at $z=h_+$ are obtained from the boundary conditions of $\mathcal{H}_x$ and $\mathcal{E}_y$ at $z=h$. Let us first define the following new quantities:

$$V^+(\eta, h) = \frac{-1}{(2\pi)^{1/2}}\int_0^\infty \mathcal{E}_y(y, h)e^{i\eta y}dy \tag{9a}$$

$$V^-(\eta, h) = \frac{-1}{(2\pi)^{1/2}}\int_{-\infty}^0 \mathcal{E}_y(y, h)e^{i\eta y}dy \tag{9b}$$

$$g^+(\eta, h) = \frac{1}{(2\pi)^{1/2}}\int_0^\infty [\mathcal{H}_x(y, h_+) - \mathcal{H}_x(y, h_-)]e^{i\eta y}dy \tag{9c}$$

$$g^-(\eta, h) = \frac{1}{(2\pi)^{1/2}}\int_{-\infty}^0 [\mathcal{H}_x(y, h_+) - \mathcal{H}_x(y, h_-)]e^{i\eta y}dy. \tag{9d}$$

It is obvious that

$$V(\eta, h_+) = V(\eta, h_-) = V^+(\eta, h) + V^-(\eta, h) \tag{10a}$$

and

$$I(\eta, h_+) - I(\eta, h_-) = 3^+(\eta, h) + g^-(\eta, h). \tag{10b}$$

Two constants will appear very often in the equations below, so for convenience we will represent them as follows:

$$a_1 = (1+s^2)^{1/2} \qquad a_1 > 0$$

$$a_2 = (1+s'^2)^{1/2} \qquad a_2 > 0.$$

From the remaining boundary conditions at $z=h$, we obtain the following results:

$$\mathcal{E}_y = 0 \qquad \text{for } y > 0, \tag{11a}$$

therefore

$$V^+(\eta, h) = 0; \tag{11b}$$

and

$$\mathcal{H}_x(y, h_+) - \mathcal{H}_x(y, h_-) = H_{0x}(y, h_-) \quad \text{for } y < 0, \tag{12a}$$

therefore

$$g^-(\eta, h) = -j \frac{\operatorname{sech}[Khs']}{(2\pi)^{1/2}[\eta + Ka_2]}, \tag{12b}$$

provided Imag $\eta <$ I mag $(-Ka_2)$.

Therefore, all the boundary conditions at $z=h$ are satisfied if

$$g^+(\eta, h) = \frac{2\omega\epsilon_0(K^2a_1^2 - \eta^2)}{(K^2 - \eta^2)^{1/2}(K^2a_2^2 - \eta^2)} G(\eta) V^-(\eta, h) + j \frac{\operatorname{sech}(Khs')}{(2\pi)^{1/2}(\eta + Ka_2)} \tag{13a}$$

where

$$G(\eta) = \frac{-j + \tan[h(K^2 - \eta^2)^{1/2}]}{2} \times \frac{K^2a_2^2 - \eta^2}{K^2a_1^2 - \eta^2} \times \frac{1 + j\dfrac{(K^2\epsilon - \eta^2)^{1/2}}{\epsilon(K^2 - \eta^2)^{1/2}} \tan[(K^2\epsilon - \eta^2)^{1/2}d]}{\tan[(K^2 - \eta^2)^{1/2}h] + \dfrac{(K^2\epsilon - \eta^2)^{1/2}}{\epsilon(K^2 - \eta^2)^{1/2}} \tan[(K^2\epsilon - \eta^2)^{1/2}d]}. \tag{13b}$$

The quantity s is the modulus of the wavenumber in the OZ direction in the air normalized with respect to K for the lowest TM surface wave in the grounded dielectric slab.

A study of the behavior of the functions in (13a) permits us to apply the Wiener-Hopf technique and solve for g^+ and V^-.

The Solution of the Equation for the Transforms

The behavior of $g^+(\eta, h)$ and $V^-(\eta, h)$ in the complex η plane is determined by the asymptotic behavior of $\mathcal{H}_x$ and $\mathcal{E}_y$ as well as by the singularities of the transformed Kernel

$$\frac{2\omega\epsilon_0(K^2a_1^2 - \eta^2)}{(K^2 - \eta^2)^{1/2}(K^2a_2^2 - \eta^2)} G(\eta). \tag{14}$$

We will come back later to (14). Let us proceed now with a physical derivation of the dominant terms of the far fields.

In our problem we can obtain all of the excited fields from the x component of the magnetic field. The problem may be compared with the two-dimensional field excited along a grounded dielectric slab by a magnetic line source along the OX axis.[8] If $-\pi/2 < \phi_0 < 0$ and ρ is very large, we will not be able to notice any difference between a magnetic line and a terminated parallel plate waveguide propagating the lowest E mode. Therefore, the nature of the solution for both problems is the same for that region of space. However, as ϕ_0 increases, the angular dependence will be different for the two problems.

In the case of the magnetic line, we would have only the surface wave for $\phi_0 \approx -\pi/2$, all other terms being of order $\rho^{-3/2}$ or lower. As ϕ_0 increases, the surface-wave contribution becomes negligible and the dominant term varies like $\rho^{-1/2}$ (always for large ρ). Finally, when ϕ_0 grows to $\pi/2$, the $\rho^{-1/2}$ terms are not present and we have only terms of order $\rho^{-3/2}$ or lower and the surface wave.

For a grounded dielectric slab excited by a parallel plate waveguide, we will have also the surface wave and terms of order $\rho^{-3/2}$ and lower if $\phi_0 \approx -\pi/2$. As ϕ_0 increases, the surface wave contribution becomes negligible and the dominant term varies like $\rho^{-1/2}$. Finally, when ϕ_0 increases to $\pi/2$, the fields will go to zero as $\rho^{-3/2}$ at least and the surface wave will not reappear if we remain outside the waveguide.

It is therefore justified to write the form of the far fields for $K\rho \gg 1$ and $\pi/2 \geq \phi_0 \geq -\pi/2$ as follows:

$$H_x = \mathcal{H}_x = g(\phi_0) \frac{e^{-jK\rho}}{\rho^{1/2}} + C\left[\frac{\cos[Kr(z+d)]}{\cos(Krd)} u(-z) + e^{-Ksz}u(z)\right]\Psi(\phi_0) \cdot \exp(jKya_1) \tag{15a}$$

$$E_y = \mathcal{E}_y = -\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2} g(\phi_0) \frac{e^{-jK\rho}}{\rho^{1/2}} \cos\phi_0 + jsC\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2}\left[\frac{\sin[Kr(z+d)]}{\sin(Krd)} u(-z) + e^{-Ksz}u(z)\right]\Psi(\phi_0) \exp(jKya_1) \tag{15b}$$

$$\Psi(\phi_0) = 0 \quad \text{if} \quad \frac{\phi_0}{2} > -\arctan(a_1 - s) \tag{15c}$$

$$\Psi(\phi_0) = 1 \quad \text{if} \quad \frac{\phi_0}{2} < -\arctan(a_1 - s). \tag{15d}$$

The transmission coefficient to the surface wave is represented by C. The new quantities ϕ_0 and ρ are the usual cylindrical coordinates illustrated in Fig. 1(a); r is the normalized wavenumber in the dielectric in the OZ direction for the lowest TM surface wave in the grounded dielectric slab. Finally, $g(\phi_0)$ is a function of the observation angle ϕ_0.

Inside the partially filled parallel plate waveguide and far away from the discontinuity, we have only the incident and reflected wave associated with the only propagating mode.

[8] The reader interested in the quantitative results for the excitation of a grounded dielectric slab by a magnetic line will find the complete discussion in Angulo and Chang, *op. cit.*, footnote 2. The physical analogy discussed in the present paragraph can easily be checked for the known results of a parallel plate waveguide radiating in free space and a magnetic line source in free space.

Therefore,

$$\mathcal{H}_x = BH_{0x} \exp\{-j2Kya_2\} \quad \text{for } Ky \gg 1 \text{ and } -d < z < h; \tag{16a}$$

$$\mathcal{E}_y = -BE_{0y} \exp\{-j2Kya_2\} \quad \text{for } Ky \gg 1 \text{ and } -d < z < h. \tag{16b}$$

B is the reflection coefficient. The quantities K, Ka_1 and Ka_2 must have small negative imaginary parts for dissipative media.

From a detailed examination of the singularities and zeros of (14) and from the asymptotic expression of the fields given in (15) and (16), it follows that $V^-(\eta, h)$ is analytic in the lower half of the η plane for Imag $\eta <$ Imag $(-Ka_1)$ and its singularities are a branch point at $\eta = -K$ and a simple pole at $\eta = -Ka_1$.

It also follows that $g^+(\eta, h)$ is analytic in the upper half of the η plane for Imag $\eta >$ Imag (Ka_2) and its singularities are a branch point at $\eta = K$, a simple pole at $\eta = Ka_2$, and a countable infinite number of poles on the negative imaginary axis $\eta = -j|\eta_i|$.[9]

In a slightly dissipative medium g^+, V^- and $G(\eta)$ are analytic in a common narrow strip of width $2w_d$ along the real axis, where

$$0 < w_d < |\text{ Imag } K|, \tag{17a}$$

$$0 < w_d < |\text{ Imag } (Ka_2)|, \tag{17b}$$

$$0 < w_d < |\text{ Imag } (Ka_1)|. \tag{17c}$$

The regions where g^+ and V^- are analytic, their singularities relevant to the integration, the overlapping strip and the branch cuts, are shown in Fig. 4. We restrict ourselves to remain on the Riemann sheet where

$$\text{Imag } (K^2 - \eta^2)^{1/2} < 0.$$

We now decompose the function G:

$$G(\eta) = \exp\{\gamma^-(\eta) - \gamma^+(\eta)\} \tag{18a}$$

where

$$\gamma^-(\eta) = \frac{-1}{2\pi j} \int_{-\infty+jw_d}^{\infty+jw_d} \frac{\ln G(\xi)}{\xi - \eta}\, d\xi \tag{18b}$$

is analytic for

$$\text{Imag } \eta < w_d$$

and

$$\gamma^+(\eta) = \frac{-1}{2\pi j} \int_{-\infty-jw_d}^{\infty-jw_d} \frac{\ln G(\xi)}{\xi - \eta}\, d\xi \tag{18c}$$

is analytic for Imag $\eta > -w_d$.

[9] $\pm j|\eta_i|$ are the roots of

$$\tan[h(K^2 - \eta^2)^{1/2}] + \frac{(K^2\epsilon - \eta^2)^{1/2}}{\epsilon(K^2 - \eta^2)^{1/2}} \tan[(K^2\epsilon - \eta^2)^{1/2} d] = 0$$

excluding

$$\eta = \pm Ka_2.$$

For the complete analysis, see Angulo and Chang, *op. cit.*, footnote 3, pp. 12–17.

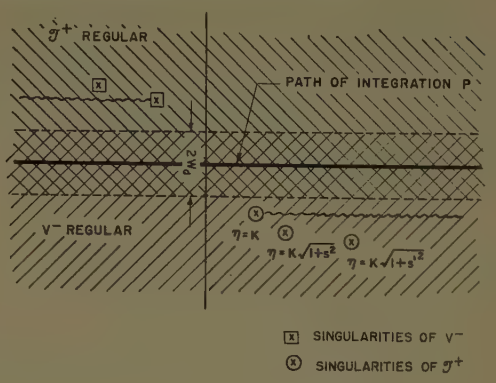


Fig. 4—The η plane.

Substituting (18a) into (13a) we can now group the terms with the same regions of regularity, as follows:

$$g^+(\eta, h) \times \frac{(K - \eta)^{1/2}(\eta - Ka_2) \exp\{\gamma^+(\eta)\}}{\eta - Ka_1} - j \frac{A \text{ sech } (Khs')}{(2\pi)^{1/2}} \times \frac{(K - \eta)^{1/2}(\eta - Ka_2)}{(\eta + Ka_2)(\eta - Ka_1)} \times \exp\{\gamma^+(\eta)\} + j \frac{2A \text{ sech } (Khs')}{(2\pi)^{1/2}} \times \frac{(K + Ka_2)^{1/2}}{\eta + Ka_2} \times \frac{a_2}{a_1 + a_2} \times \exp\{\gamma^+(-Ka_2)\}$$

$$= \frac{2\omega\epsilon_0}{(K + \eta)^{1/2}} \times \frac{\eta + Ka_1}{\eta + Ka_2} \times V^-(\eta, h) \times \exp\{\gamma^-(\eta)\} + j \frac{2A \text{ sech } (Khs')}{(2\pi)^{1/2}} \times \frac{(K + Ka_2)^{1/2}}{\eta + Ka_2} \times \frac{a_2}{a_1 + a_2} \times \exp\{\gamma^+(-Ka_2)\}.$$

The last terms on each side of the previous equation are identical. They have been added in order to eliminate the pole at $\eta = -Ka_2$ of the left hand side of the equation.

The asymptotic behavior $\eta \to \infty$ of the unknown functions g^+ and V^- in their respective regions of regularity is given by the behavior of the unknown field around the edge.[10] The physical requirement that the field scattered by the edge must have a finite amount of energy requires

$$\frac{V^-(\eta, h)}{\sqrt{\eta}} \to 0 \quad \text{and} \quad \sqrt{\eta}\, g^+(\eta, h) \to 0$$

as $\eta \to \infty$ in the region of regularity. These conditions are identical to those for the scattering of a plane wave by half a plane.[11] We can now proceed with the customary reasoning of the Wiener-Hopf technique. The left hand side of the above equation is analytic in the upper half of the η plane including the narrow strip $|\text{I mag } \eta| < w_d$ and the right hand side is analytic in the lower half including the narrow strip. Therefore, they must be an-

[10] Morse and Feshbach, *op. cit.*, p. 462.

[11] E. T. Copson, *Quarterly J. of Math. (Oxford)*, vol. 17, no. 65, pp. 19–34; March, 1946.

alytic continuation of each other representing an entire function of η. Furthermore, both sides approach zero as η approaches infinity. Thus, the entire function is zero. Equating both sides to zero we obtain[12]

$$g^+(\eta, h) = \frac{j \text{ sech } (Khs')}{(2\pi)^{1/2}(\eta + Ka_2)} \times \left\{1 - \frac{2a_2[K + Ka_2]^{1/2}(\eta - Ka_1)}{(K - \eta)^{1/2}(a_2 + a_1)(\eta - Ka_2)} \exp\ [\gamma^+(-Ka_2) - \gamma^+(\eta)]\right\} \quad (19a)$$

and

$$V^-(\eta, h) = \frac{-j \text{ sech } (Khs')}{\omega\epsilon_0 (2\pi)^{1/2}(\eta + Ka_1)} \times \frac{(K + \eta)^{1/2}a_2(K + Ka_2)^{1/2}}{a_2 + a_1} \exp \cdot \{\gamma^+(-Ka_2) - \gamma^-(\eta)\}. \quad (19b)$$

Calculation of the Far Fields

The "voltage" is now known for $z = h$, since

$$V(\eta, h) = V^-(\eta, h). \quad (20a)$$

As for the "current" transform, we recall (7b) for $z = h$:

$$I(\eta, h_+) = Y_a V(\eta, h). \quad (20b)$$

Therefore $I(\eta, h_+)$ is known. Moreover, (10b), (12b) and (19a) yield $I(\eta, h_-)$, once $I(\eta, h_+)$ has been found. Therefore, $I(\eta, h_+)$ and $I(\eta, h_-)$ are now both known.

The knowledge of V and I for $z = h$ gives us the expressions for V and I anywhere, as indicated in (7) and (8).

Finally, the inversion of V and I by (5) yields the exact expressions for the fields everywhere.

However, this solution everywhere is only formal, since the inversion of the transforms is practically impossible. Nevertheless, we can obtain all the information that we want by carrying out the inversion for the far fields at points of observation for which the method of steepest descents is easier to apply, and comparing the results with (15) and (16). In this way we can obtain the expressions for the coefficients B, C, and $g(\phi_0)$ which give us the complete knowledge of the far fields.

The inverse transforms of special interest to us are:

$$E_y(y, h) = \mathcal{E}_y(y, h) = \frac{-1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} V^-(\eta, h) e^{-iyv} d\eta \quad (21a)$$

for $z = h$, $y < 0$,

$$H_x(y, h_+) - H_x(y, h_-) + H_{0x}(y, h) = \mathcal{H}_x(y, h_+) - \mathcal{H}_x(y, h_-) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty P}^{+\infty} \left[g^+(\eta, h) - \frac{j \text{ sech } (Khs')}{(2\pi)^{1/2}(\eta + Ka_2)}\right] e^{-iyv} d\eta \quad (21b)$$

for $z = h$, $y > 0$, and

$$E_y(y, z) = \mathcal{E}_y(y, z) = \frac{-1}{(2\pi)^{1/2}} \int_{-\infty P}^{\infty} V^-(\eta, h) e^{-iyv} \exp\left\{-j(K^2 - \eta^2)^{1/2}(z - h)\right\} d\eta \quad \text{for } z > h. \quad (21c)$$

The path of integration is indicated in Fig. 4.

The integrals are evaluated for the far fields by the method of steepest descents for the limiting case of zero dissipation. For convenience these integrations are not carried out in the η plane but in a new ν plane. The coordinates are also changed from cartesian to polar coordinates, as illustrated in Fig. 1(a).

$$\eta = K \sin \nu \qquad (K^2 - \eta^2)^{1/2} = K \cos \nu \quad (22a)$$

$$y = \rho \sin \phi_0 \qquad z - h = \rho \cos \phi_0. \quad (22b)$$

With this change, the integrals (21) become

$$E_y(y, h) = \frac{-K}{(2\pi)^{1/2}} \int_A V^-(K \sin \nu, h) \cdot \exp\left\{-jK\rho \cos\left(\nu + \frac{\pi}{2}\right)\right\} \cos \nu d\nu \quad (23a)$$

for $z = h$, $y < 0$,

$$H_x(y, h_+) - H_x(y, h_-) + H_{0x}(y, h) = \frac{K}{(2\pi)^{1/2}} \int_A \left[g^+(K \sin \nu, h) - j \frac{\text{sech } (Khs')}{K(2\pi)^{1/2}(\sin \nu + a_2)}\right] \cdot \exp\left\{-K\rho j \cos\left(\nu - \frac{\pi}{2}\right)\right\} d\nu \quad (23b)$$

for $z = h$, $y > 0$, and

$$E_y(y, z) = \frac{-K}{(2\pi)^{1/2}} \int_A V^-(K \sin \nu, h) \cdot \exp\left\{-jK\rho \cos(\nu - \phi_0)\right\} \cos \nu d\nu \quad (23c)$$

for $z > h$.

The path of integration is shown in Fig. 5.

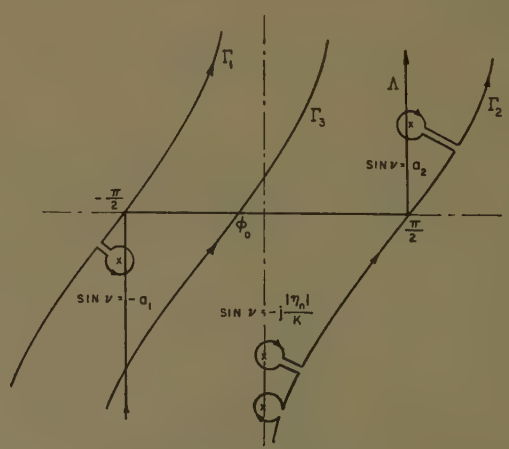


Fig. 5—The ν plane.

[12] For full algebraic analysis see Angulo and Chang, *op. cit.*, footnote 3, pp. 19–21.

The saddle point for (23a) is obviously at $\nu = -\pi/2$, and the steepest descents path is the path Γ_1 in Fig. 5. The integration indicated in (23a) is carried out along the path Γ_1 and not along the path Λ. The result is equal to $2\pi j$ times the residue of the integrand at the pole $\sin\nu = -a_1$ plus the asymptotic series obtained from the steepest descend integration. However, because of the factor $\cos\nu$, the $\rho^{-1/2}$ order contribution of the expansion is zero since $\cos(-\pi/2) = 0$. Therefore, if we neglect terms of order $\rho^{-3/2}$ or lower we can write for large ρ, (*i.e.*, $y \ll 0$)

$$E_y(y, h) = -(2\pi)^{1/2} j e^{jKa_1 y} \lim_{\eta \to -Ka_1} \{(\eta + Ka_1) V^-(\eta, h)\}. \quad (24a)$$

This represents the surface wave excited along the grounded slab.

If we evaluate (15b) for the given observation point

$$z = h \qquad \phi_0 = -\frac{\pi}{2}$$

$$y \ll 0 \qquad \rho \text{ very large,}$$

we obtain

$$E_y(y, h) = jsC\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2} e^{-Ksh} e^{jKa_1 y}. \quad (24b)$$

Equating the right hand members (24), one obtains:

$$C = -\frac{(2\pi)^{1/2}}{s}\left(\frac{\epsilon_0}{\mu_0}\right)^{1/2} \lim_{\eta \to -Ka_1} \{(\eta + Ka_1) V^-(\eta, h)\} e^{Ksh}$$

The evaluation of the limit yields finally:

$$C = j\frac{a_2(1 - a_1)^{1/2}(1 + a_2)^{1/2}}{s(a_1 + a_2)} \times \operatorname{sech}(Khs') \times \exp\{Khs + \gamma^+(-Ka_2) - \gamma^-(-Ka_1)\}. \quad (25)$$

The saddle point for (23b) is at $\nu = \pi/2$ and the steepest descent path is Γ_2 in Fig. 5. The residues at the poles on the imaginary axis of the ν plane decay exponentially with an increasing positive y and are negligible for $Ky \gg 1$. These poles correspond to the nonpropagating ordinary modes in the parallel plate waveguide with the dielectric slab, and we expected a negligible contribution for $Ky \gg 1$. The asymptotic expansion does not contain a term of order $\rho^{-1/2}$ for identical reasons that the integral discussed in the previous paragraph did not. The dominant contribution to the integral (23b) is therefore:

$$\mathcal{H}_x(y, h_+) - \mathcal{H}_x(y, h_-) = -(2\pi)^{1/2} j e^{-jKa_2 y} \times \lim_{\eta \to Ka_2} \left\{(\eta - Ka_2)\left[g^+(\eta, h) - \frac{j \operatorname{sech}(Khs')}{(2\pi)^{1/2}(\eta + Ka_2)}\right]\right\} \quad (26a)$$

where only terms of order $\rho^{-3/2}$ or lower have been neglected and $Ky \gg 1$. The expression (26a) clearly represents the reflected surface wave in the parallel plate waveguide evaluated at the upper plate.

If we evaluate (15a) and (16a) at the point of observation

$$z = h \qquad \phi_0 = \frac{\pi}{2}$$

$$Ky = 1 \qquad \rho \text{ very large,}$$

and subtract, we obtain

$$\mathcal{H}_x(y, h_+) - \mathcal{H}_x(y, h_-) = g\left(\frac{\pi}{2}\right)\frac{e^{-jK\rho}}{\rho^{1/2}} - B\frac{e^{-jKa_2 y}}{\cosh(Khs')}. \quad (26b)$$

The identity of (26a) and (26b) requires

$$g\left(\frac{\pi}{2}\right) = 0$$

and

$$B = (2\pi)^{1/2} j \cosh(Khs') \lim_{\eta \to Ka_2} \cdot \left\{(\eta - Ka_2)\left[g^+(\eta, h) - \frac{j \operatorname{sech}(Khs')}{(2\pi)^{1/2}(\eta + Ka_2)}\right]\right\}. \quad (26c)$$

Finally

$$B = \frac{(1 + a_2)^{1/2}(a_2 - a_1)}{(1 - a_2)^{1/2}(a_2 + a_1)} \exp\{\gamma^+(-Ka_2) - \gamma^-(Ka_1)\}. \quad (27)$$

The evaluation of the integral (23c) is carried out in the same way as with the two discussed above. The saddle point is at $\nu = \phi_0$ and the path of integration is deformed now from Λ to Γ_3 in Fig. 5.

The observation angle may vary $-\pi/2 < \phi_0 < \pi/2$; and depending on the value of ϕ_0, we will have or will not have a residue contribution from the pole at $\sin\nu = -a_1$. This contribution is the surface wave excited on the grounded dielectric slab. The asymptotic series will contribute this time with a term of the order $\rho^{-1/2}$ provided $|\phi_0| < \pi/2$. The dominant term of the expansion plus the pole contribution yield

$$E_y(y, z) = -(1 + j)\left(\frac{K}{2\rho}\right)^{1/2} e^{-jK\rho} \cos\phi_0 V^-(K \sin\phi_0, h) - j(2\pi)^{1/2} \exp\{jKa_1 y - sK(z - h)\} \times u\left[-\frac{\phi_0}{2} - \tan^{-1}(a_1 - s)\right] \times \lim_{\eta \to -Ka_1} \{(\eta + Ka_1) V^-(\eta, h)\} \quad (28a)$$

where $z > h$, $K\rho \gg 1$, $-\pi/2 < \phi_0 < \pi/2$, and $u(x)$ is the Heaviside unit step function. If the observation angle is $\phi_0 > -2\tan^{-1}(a_1 - s)$, *i.e.*, if we are in the region with no surface wave:

$$E_v(y, z) = -(1+j)\left(\frac{K}{2\rho}\right)^{1/2} e^{-jK\rho}\cos\phi_0 V^-(K\sin\phi_0, h). \quad (28b)$$

We can now evaluate (15b) for

$$z > 0 \qquad \phi_0 > -2\tan^{-1}(a_1 - s)$$
$$\rho \text{ very large}$$

and obtain:

$$E_v(y, z) = -\left(\frac{\mu_0}{\epsilon_0}\right)^{1/2} g(\phi_0)\cos\phi_0 \frac{e^{-jK\rho}}{\rho^{1/2}}. \quad (28c)$$

The identity of (28b) and (28c) requires

$$g(\phi_0) = (1+j)\frac{\omega\epsilon_0}{(2K)^{1/2}} V^-(K\sin\phi_0, h).$$

Replacing V^- by its value, one obtains finally

$$g(\phi_0) = \frac{(1-j)(1+\sin\phi_0)^{1/2} a_2(1+a_2)^{1/2}\operatorname{sech}(Khs')}{2\pi(K)^{1/2}(\sin\phi_0 + a_1)(a_2 + a_1)} \times \exp\{\gamma^+(-Ka_2) - \gamma^-(K\sin\phi_0)\}. \quad (29)$$

in the physics of the problem.[13]

The component E_z is not evaluated explicitly. It is related simply to H_x through the relation

$$E_z = \frac{j}{\omega\epsilon_0\epsilon(z)}\frac{\partial H_x}{\partial y}. \quad (30)$$

Because of the orthogonality of the modes of either the grounded dielectric slab or the partially filled parallel plate waveguide, the powers carried by the launched surface wave and the radiated power are not coupled to each other. The reflected power, the radiated power, and the power in the surface wave per unit incident power can be found individually from Poynting's theorem.

The percentage of power reflected is:

$$P_{ref} = \frac{(1+a_2)^2(a_2 - a_1)^2}{(a_2 - 1)^2(a_2 + a_1)^2} \times \exp\left\{\frac{-4a_2}{\pi}\int_0^1 \frac{\Omega(x)}{1 + s'^2 - x^2}\,dx\right\}. \quad (31a)$$

The percentage of power transmitted is:

$$P_{trans.} = \frac{4(a_1 - 1)a_2(a_2 + 1)a_1\operatorname{sech}^2(Khs')}{s(a_2 - 1)(a_2 + a_1)^2[1 + \tanh(Khs)]^2} \times \exp\left\{2Khs - \frac{a_2}{2a_1}\ln\frac{a_1 + 1}{a_1 - 1} - \frac{2}{\pi}\int_0^1\left(\frac{a_2}{a_2^2 - x^2} - \frac{a_1}{a_1^2 - x^2}\right)\Omega(x)dx\right\} \quad (31b)$$

$$\Omega(x) = Kh(1 - x^2)^{1/2} + \arctan\left\{\frac{(\epsilon - x^2)^{1/2}}{\epsilon(1 - x^2)^{1/2}}\tan\left[Kd(\epsilon - x^2)^{1/2}\right]\right\}. \quad (31c)$$

The percentage of power radiated per unit angle in the direction ϕ_0 (see Fig. 1) is:

$$P_{rad.} = \frac{2a_2(a_2 + 1)(s'^2 - s^2)(1 + \sin\phi_0)(s^2 + \cos^2\phi_0)}{\pi(a_1 + \sin\phi_0)^2(a_2 + a_1)^2(a_2 - 1)} \times \frac{\left|\tan(Kh\cos\phi_0) + \dfrac{(\epsilon - \sin^2\phi_0)^{1/2}}{\epsilon\cos\phi_0}\tan\left[Kd(\epsilon - \sin^2\phi_0)^{1/2}\right]\right|}{|\operatorname{sech}(Kh\cos\phi_0)|\,(s'^2 + \cos^2\phi_0)\left\{1 + \dfrac{\epsilon - \sin^2\phi_0}{\epsilon^2\cos^2\phi_0}\tan^2\left[Kd(\epsilon - \sin^2\phi_0)^{1/2}\right]\right\}^{1/2}} \times \exp\left\{\frac{-2a_2}{\pi}\int_0^1 \frac{\Omega(x)dx}{a_2^2 - x^2} + \frac{2\sin\phi_0}{\pi}\mathcal{P}\int_0^1 \frac{\frac{\pi}{2} - \Omega(x)}{\sin^2\phi_0 - x^2}\,dx\right\}. \quad (31d)$$

With the knowledge of B, C and $g(\phi_0)$, we can complete (15) and (16). Therefore, we have the far fields everywhere. The computation of γ^+ and γ^- involves integration in a complex plane and careful definitions of the paths of integration, but there are no serious difficulties and the integrations have no special importance

Symbol $\mathcal{P}$ in (31d) stands for Cauchy's Principal Value.

The values of (31) are calculated for $\epsilon = 2.49$. The integrals in the exponents are calculated numerically. Figs. 6, 7, and 8 are plots of the power reflected, the

[13] For their evaluation, see Angulo and Chang, *op. cit.*, footnote 3.

Fig. 6—Percentage of power transmitted to the surface wave vs thickness of the grounded slab for various heights of the parallel plate waveguide.

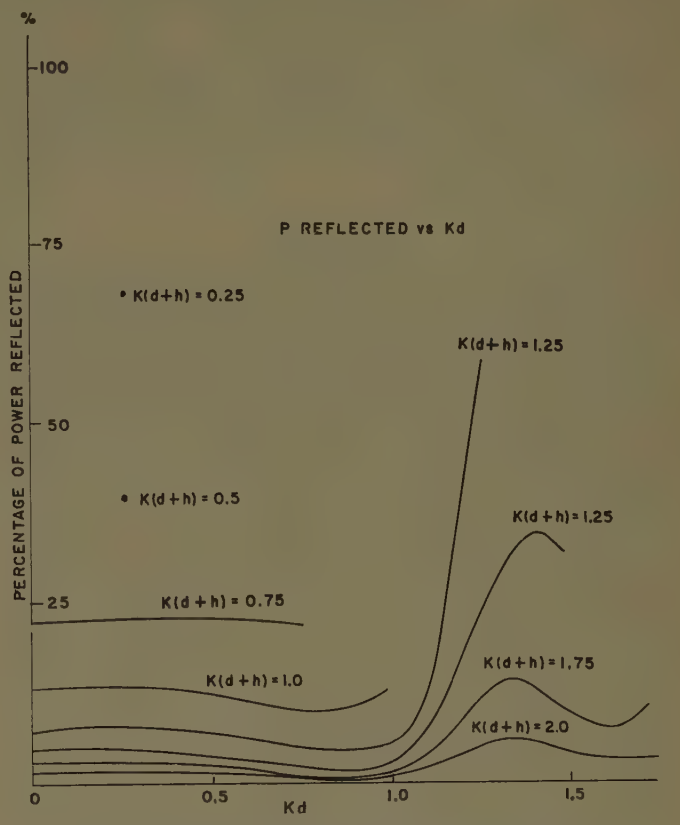


Fig. 7—Percentage of power reflected vs thickness of the grounded slab for various heights of the parallel plate waveguide.

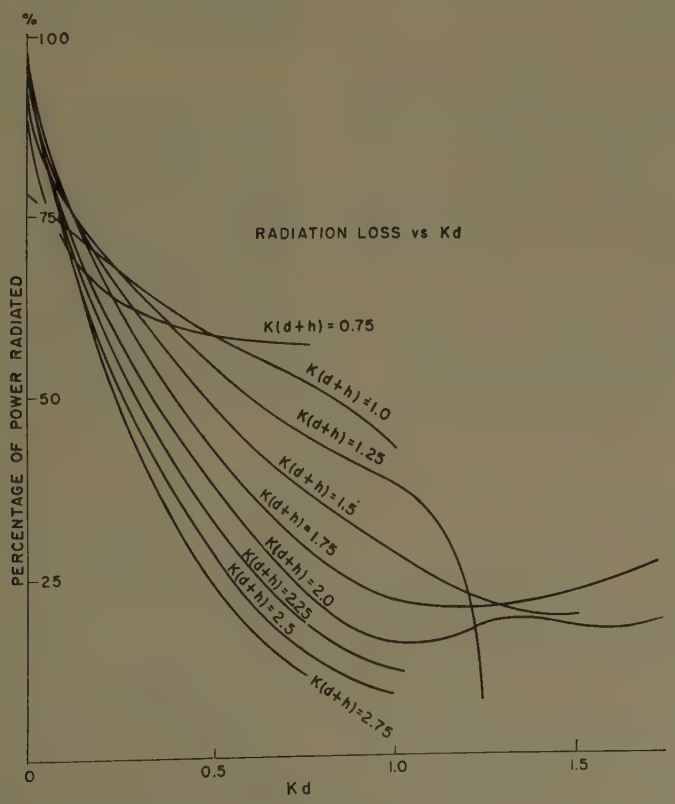


Fig. 8—Percentage of total power radiated vs thickness of the grounded slab for various heights of the parallel plate waveguide.

power transmitted and the total power radiated vs the normalized thickness of the slab, Kd, for various heights of the parallel plate waveguide, $K(d+h)$. Notice that most curves are terminated at the points for which the guide is completely filled with dielectric, *i.e.*, $d+h=d$ or $h=0$ (see Fig. 1).

Conclusions

Fig. 3 shows that s' is smaller for thin dielectric slabs and therefore the incident surface wave decays slowly when one moves away from the slab toward the upper plate inside the parallel plate waveguide. This agrees with the physical expectation that the larger part of the incident energy will be in the air region of the parallel plate waveguide if the dielectric slab is thin. Under those circumstances, the radiation produced by the discontinuity in the upper plate should be very pronounced. This indicates that the percentage of power transmitted to the surface wave on the grounded dielectric slab should grow with d, and if d is fixed should grow by separating further the upper plate (that is, increasing h). One can see the verification of this reasoning in the results shown in Fig. 6. It is not surprising to see that the total radiated power has a relative maximum if h is varied, while d remains constant and not large. This is shown in the crossing of the curves of Fig. 8 for small d.

If the waveguide is almost completely filled with dielectric, the effect of the upper plate will again influence strongly the radiation because of its proximity to the slab. This explains the crossing of the curves in Fig. 8 for large d, which gives a relative maximum of total radiation for a given h and d fixed.

The efficiency of excitation of the grounded dielectric slab is large if we choose $K(d+h)$ and Kd large enough, as indicated in Fig. 6; but we must keep in mind the limitation in $K(d+h)$ if only one mode is going to propagate in the parallel plate waveguide containing the slab. It is interesting to note that the maximum efficiency for all $K(d+h)\geq 1$ occurs at $Kd\simeq 1.0$.

From the point of view of the reflection, Fig. 7 indicates that it is fairly easy to choose the parameters in order to have an almost perfect match.

The excitation would be very practical and efficient, for instance, for $K(d+h)=2.0$ and $0.75<Kd<1.25$. In this range of d we have a very small amount of power reflected (Fig. 7) and the percentage of power radiated (Fig. 8) as well as the percentage of power into the infinite grounded dielectric slab will not vary critically with Kd.

It is instructive to compare this means of exciting an infinite grounded dielectric slab with the known results for the excitation by a slot[2,14,15] by an electric dipole.[15–17]

The magnetic line and the dipole provide excitation efficiencies as high as 95 per cent, but this maximum efficiency drops fast if the source is not in the optimum location. The maximum efficiency of excitation by a parallel plate waveguide does not present a critical dependence on the parameters if these are chosen in the right range (Fig. 6).

Another advantage of the waveguide above the singular sources is that for the same excitation efficiency the undesired radiation caused by the discontinuity of the excitation device is smaller for the parallel plate waveguide, because part of the power which is not transmitted is reflected back into the waveguide. This effect is particularly obvious when $Kd>1$ for all $K(d+h)$, as we can see in Fig. 8.

Finally, it should be emphasized here that the analytic answer given for the amplitudes of the transmitted and reflected waves are exact. The only approximation which has been made in this paper is the evaluation of $g(\phi_0)$ where only the first term of the asymptotic series is used. However, it can easily be improved by evaluating more terms of the asymptotic series in the integration by the method of steepest descents.

[14] A. L. Cullen, "The excitation of plane surface waves," *Proc. IEE (London)*, vol. 101, pt. IV, no. 7, pp. 225–230; August, 1954.
[15] B. Friedman and W. E. Williams, "Excitation of Surface Waves," Div. of Electromagnetic Res., Inst. of Math. Sciences, New York Univ., Res. Rept. No. EM-99; October, 1956. AFCRC-TN-56-798, ASTIA Doc. No. AD 119139.
[16] C. M. Angulo and W. S. Chang, "On the Excitation of the Surface Waves in Dielectric Slabs." pt. III, Rept. No. AF 1391/4 Div. of Engrg., Brown Univ., Providence, R. I.; AFCRC-TN-56-952, ASTIA Doc. No. AD 110143; October, 1956.
[17] D. Brick, "The Radiation of a Hertzian Dipole Over a Coated Conductor," Cruft Lab., Harvard Univ., Cambridge, Mass., Tech. Rept. No. 172; 1953.

Random Errors in Aperture Distributions*

R. H. T. BATES†

Summary—**The effects of random manufacturing errors on polar diagrams of antennas are analyzed in terms of the radius of correlation and mean square magnitude of the errors. The basis of the method is the Wiener-Khintchine theorem. Approximate general formulas are given for the reduction in gain and lowest probable side-lobe level. The implications of the theory are discussed.**

Introduction

THERE have been many accounts of the effects of random manufacturing errors on antenna polar diagrams. The majority of these consider aperture distributions made up of discrete radiators.[1,2] Robieux[3] and Ruze[4] treat continuous distributions. The analysis applied here is similar to theirs, but arrives at the result rather more directly, perhaps because of the use made of the Wiener-Khintchine theorem.[5]

The purpose of this note is to present a simple theory of the effect of random manufacturing errors on antenna polar diagrams. The errors are considered as superimposing random noise on the desired aperture distribution. The autocorrelation function of the errors is used in the analysis rather than the errors themselves. This follows Wiener's work on the filtering of noisy signals. The advantage of autocorrelation functions is that they are usually well behaved and recognizable, which is not true of the random processes from which they are derived.

Very approximate formulas are derived which provide quick estimates of the loss of gain and minimum side-lobe level likely to result from a certain mean manufacturing error. The formulas demonstrate that the average extent or radius of correlation of the errors is more important than their magnitude. The major effect of the errors is to raise the side-lobe level. Reduction in gain is less noticeable.

The theory, which is developed in the next section, suggests certain design and manufacturing techniques for antennas; these are discussed briefly at the end of this note.

* Manuscript received by the PGAP, September 22, 1958. Revised manuscript received by the PGAP, April 18, 1959.

† Can. Westinghouse Co., Ltd., Hamilton, Ontario, Can.

[1] E. N. Gilbert and S. P. Morgan, "Optimum design of directive antenna arrays subject to random variations," *Bell Syst. Tech. J.*, vol. 34, pp. 637–663; May, 1955.

[2] R. S. Elliott, "Mechanical and electrical tolerances for two-dimensional scanning antenna arrays," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-6, pp. 114–120; January, 1958.

[3] J. Robieux, "Influence de la precision de fabrication d'une antenne sur ses performances," *Ann. Radioelect*, pp. 29–56; January, 1956.

[4] J. Ruze, "The effect of aperture errors on the antenna radiation pattern," *Nuovo Cimento*, vol. 9, Suppl. No. 3, pp. 364–380; 1952.

[5] D. A. Bell, "Statistical Methods in Electrical Engineering," Chapman and Hall, Ltd., London, Eng., p. 89; eq. 6.22; 1953.

Theory

In order that the analysis shall be as simple as possible, the radiating aperture of the antenna is taken to lie in a plane normal to the direction of the main beam. Although very few antennas are actually flat, this procedure is well accepted.[6]

Fig. 1 represents the aperture plane of an antenna.

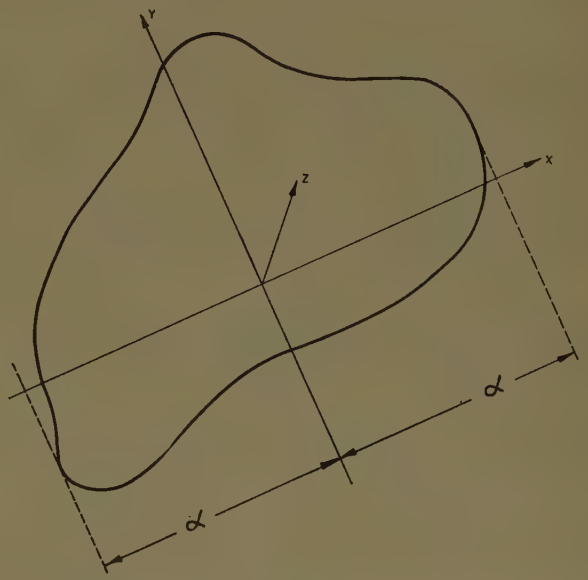


Fig. 1—Coordinates of aperture plane; z points into the paper.

The origin of coordinates is chosen so that the aperture extends equally on both sides of the y-axis in the $x-z$ plane. Let

$i(x, y)$ = aperture distribution,
$I(x)$ = equivalent distribution for calculating polar diagram in the $x-z$ plane, and
α = half length of aperture in $x-z$ plane,

where

$$I(x) = \int_{\text{aperture}} i(x, y)dy. \tag{1}$$

Fig. 2 gives the coordinate system for calculating the far-field polar diagram in the $x-z$ plane. Let

$E(\psi)$ = voltage or current far-field polar diagram,
$\psi = (\sin\theta)/\lambda$,
λ = wavelength of radiation.

[6] S. Silver, "Microwave Antenna Theory and Design," McGraw-Hill Book Co., Inc., New York, N. Y., ch. 6; 1949.

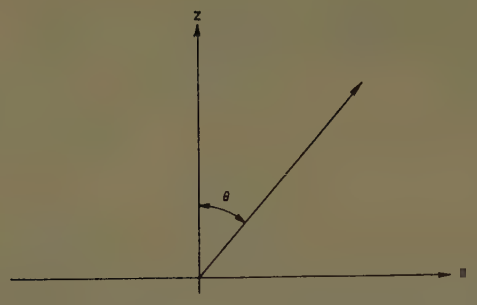


Fig. 2—Coordinates for calculating polar diagram.

Then[7]

$$E(\psi) = \int_{-\alpha}^{\alpha} I(x) e^{j2\pi x\psi} dx. \tag{2}$$

Define the auto-correlation function $\phi(\rho)$ of $I(x)$ as

$$\begin{aligned}\phi(\rho) &= \frac{1}{2\alpha}\int_{-\alpha}^{\alpha-\rho} I(x)\overline{I(x+\rho)}dx; \qquad 0 \le \rho \le 2\alpha \\ &= \frac{1}{2\alpha}\int_{-(\alpha+\rho)}^{\alpha} I(x)\overline{I(x+\rho)}dx; \quad -2\alpha \le \rho \le 0 \\ &= 0; \qquad |\rho| > 2\alpha, \end{aligned} \tag{3}$$

where $\overline{I(x)}$ = complex conjugate of $I(x)$.

Let

$$P(\psi) = E(\psi)\overline{E(\psi)} = \text{power polar diagram.} \tag{4}$$

The Wiener-Khintchine theorem[5] states that

$$P(\psi) = 4\alpha \int_0^{\infty} \phi(\rho) \cos 2\pi\rho\psi d\rho. \tag{5}$$

The manufacturing errors are taken to superimpose noise on the desired aperture distribution.

$$I(x) = L(x) + H(x) \tag{6}$$

where

$L(x)$ = desired aperture distribution in $x-z$ plane.
$H(x)$ = random function due to errors in aperture distribution.

$I(x)$ has been derived from the aperture distribution $i(x, y)$. By analogy with (5) and (1)

$$i(x, y) = l(x, y) + h(x, y),$$

where

$l(x, y)$ = desired distribution,
$h(x, y)$ = random function due to errors in aperture distribution, and

$$H(x) = \int_{\text{aperture}} h(x, y)dy. \tag{7}$$

[7] T. T. Taylor, "Design of line source antennas for narrow beamwidth and low side lobes," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-3, p. 17; January, 1955.

It is necessary to normalize $L(x)$ in order to obtain quantitative results. The most convenient way seems to be to make the average value of $L(x)$ equal to unity. Write

$$L(x) = 1 + \sum_{n=1}^{\infty}\left[A_n \cos\frac{n\pi x}{\alpha} + B_n \sin\frac{n\pi x}{\alpha}\right]. \tag{8}$$

Eq. (8) is a general expression for $L(x)$. It has the useful property that the expression for the polar diagram in the direction of the main beam does not contain any of the A_n or B_n. Let

$E_L(\psi)$ = desired voltage or current polar diagram,
$P_L(\psi)$ = desired power polar diagram, and substitute (8) into (2) and use (4).

$$E_L(0) = 2\alpha, \tag{9}$$

$$P_L(0) = 4\alpha^2. \tag{10}$$

Let

$E_H(\psi)$ = voltage or current polar diagram due to errors in aperture distribution, and
$P_H(\psi)$ = power polar diagram due to errors in aperture distribution.

Since the noise due to manufacturing errors has been assumed to be random, $E_L(\psi)$ and $E_H(\psi)$ can be taken as statistically independent. This means that the most likely value for the combined polar diagram will be obtained by adding the squares of $E_L(\psi)$ and $E_H(\psi)$. Thus the side-lobe level is unlikely ever to be less than $P_H(\psi)/P_L(0)$. Using (10),

$$\begin{aligned} S(\psi) &= \text{lowest probable side lobe in the direction } \psi \\ &= \frac{P_H(\psi)}{4\alpha^2}, \end{aligned} \tag{11}$$

provided $S(\psi) > P_L(\psi)$.

The need for the above condition is defined by the fact that the noise power in the error polar diagram will not be noticeable if it is less than the level of the desired polar diagram (produced by an aperture free from errors).

The power in the noise polar diagram is liable to be spread over a larger angular area than the main beam of the desired polar diagram; thus the gain of an antenna is given by the amplitude of the main beam divided by the sum of the radiated powers of the errors and the desired aperture distribution.

$$\begin{aligned} \text{Gain reduction factor} &= R \\ &= 1 - \frac{\text{gain in presence of errors}}{\text{gain without errors}}, \\ &\doteq \frac{\phi_H(0)}{P}, \end{aligned} \tag{12}$$

provided the power in the errors is at least an order less than that in the desired distribution.

P = total power radiated from desired distribution.

Approximate General Expressions for S and R

In this section general expressions for $S(\psi)$ and R are derived. They are necessarily approximate since they are supposed to apply to any antenna.

It is necessary to find some simple form for the autocorrelation function of the errors. Its Fourier transform must be wholly positive since it is a power polar diagram (5). A suitable choice seems to be exp $(-\rho^2/\beta^2)$. Provided β is somewhat less than α, then $\phi(\rho)$ can be taken to exist from $\rho = -\infty$ to $\rho = +\infty$. If (3) is examined it is seen that $\phi(0)$ is the mean square value of $I(x)$; this allows a magnitude to be given to $\phi(\rho)$.

Let M = mean square value of $H(x)$ if $L(x) \equiv 1$.

M is solely dependent on the mechanical errors and could be computed exactly, provided the exact form of the errors were known. It is a measure of the simple mechanical tolerance set on the construction of the antenna and is largely independent of the radius of correlation.

The actual mean square value of $H(x)$ is M multiplied by a weighting factor dependent upon the form of the desired aperture distribution $L(x)$. $H(x)$ at any point is proportional to $L(x)$ at that point. Thus, MW = mean square value of $H(x)$, where

$$W = \text{weighting factor due to form of } L(x)$$

$$= \frac{1}{2\alpha}\int_{-\alpha}^{\alpha} L(x)\ \overline{L(x)}\ dx.$$

From (8)

$$W = 1 + \frac{1}{2}\sum_{n=1}^{\infty}[A_n\overline{A}_n + B_n\overline{B}_n]. \tag{13}$$

The approximate general expression for the autocorrelation function of $H(x)$ can now be written as

$$\phi_H(\rho) = MWe^{-\rho^2/\beta^2}, \tag{14}$$

where

β = radius of correlation within which 85 per cent of a particular error is located.

Substitute (14) into (5); then

$$P_H(\psi) = 2\pi^{1/2}\alpha MW\beta e^{-\pi^2\beta^2\psi^2}. \tag{15}$$

$P_H(\psi)$ is the power polar diagram of the errors in the aperture distribution. Owing to the simple expression (14), taken for $\phi_H(\rho)$, the peak of $P_H(\psi)$ coincides with the main beam of the desired polar diagram. It is suggested that (15) should be interpreted to mean that it is likely that there will be spurious side lobes of the level $P_H(\psi)/P_H(0)$ somewhere in the polar diagram. Also, the angular area covered by these side lobes will be similar to the effective width of $P_H(\psi)$.

Examine (15). The amplitude of $P_H(\psi)$ is proportional to β, the radius of correlation. This is the practical significance of the theory. Its more obvious consequences will be discussed in the next section.

Substitute (15) into (11) and (12); then

$$S(\psi) = \frac{\pi^{1/2}\beta MW}{2^{1/2}\alpha}e^{-\pi^2\beta^2\psi^2}, \tag{16}$$

$$R = \frac{\phi_H(0)}{P} = \frac{MW}{P}. \tag{17}$$

Consequences of the Theory

The main consequence of the theory is that the mechanical design of antennas should center upon keeping the areas of influence of particular errors as small as possible. For instance, large reflectors should be supported at many points, and each point should be independently set up. Any process, such as welding, which is liable to put strains on the whole structure, should be completed before the surface of the reflector is finally positioned. Since $S(\psi)$, from (16), is proportional to β, an error stretching most of the length of the antenna is liable to have a worse effect than a localized bump or dent of much greater amplitude.

The theoretical approach used in this note suggests a checking procedure for antennas while they are being manufactured. To illustrate the procedure, it is supposed that a reflector is being built. It is necessary for the reflecting surface to be set up against a template. It is suggested that instead of merely seeing that the surface is within some specified tolerance, the positional error should be recorded at each point which is checked. The oval curve in Fig. 3 represents the projection of the reflector on some plane perpendicular to the direction of the main beam (the aperture plane). The intersections of the dotted lines are the points checked. There should be enough of these points to include all appreciable undulations of the reflector surface. Suppose that the polar diagram in the x-z plane is the one being considered.

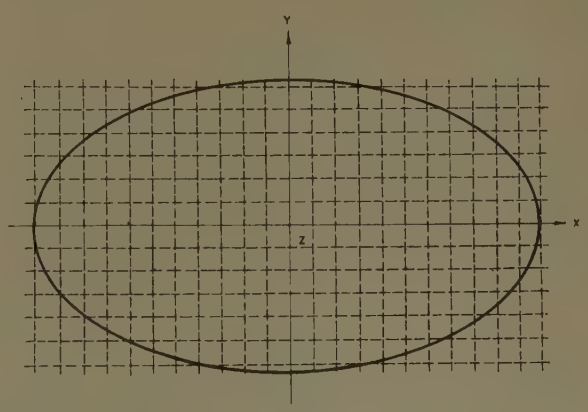


Fig. 3—Orthogonal projection of reflector on aperture plane and fixed points.

The errors on any vertical dotted line are averaged. This gives a series of averaged errors at points along the x-axis separated by the distance between adjacent dotted lines. The mechanical error is converted into aperture distribution error using the known variation of the aperture distribution $L(x)$, with x. Let

$2m+1$ = number of points on x-axis,
ϵ = separation of adjacent points,
H_P = error in aperture distribution at point P; and
ϕ_q = autocorrelation function of errors.

ϕ_q will consist of a discrete set of values since the errors have been given as such.

$$\phi_q = \frac{\epsilon}{2m} \sum_{p=-m}^{m-q} H_p \overline{H_{p+q}}; \qquad 0 \leq q \leq 2m$$

$$= \frac{\epsilon}{2m} \sum_{p=-(m+q)}^{m} H_p \overline{H_{p+q}}; \quad -2m \leq q \leq 0$$

$$= 0; \qquad 2m < |q| < \infty. \tag{18}$$

The computations for deriving the H_P and ϕ_q are ideally suited to digital computation. The programming is simple and the machine time would be only a few minutes even for large antennas. It is worth noting that only the first part of (18) need be calculated since autocorrelation functions are necessarily even. ϕ_q is converted to a continuous function by drawing a smooth curve through the discrete points.

The magnitude and shape of ϕ_q are sufficient for estimating the effect of the errors on the polar diagram. $2m\phi_0$ is the power in the error polar diagram and the width of ϕ_q is roughly inversely proportional to the width of the main beam of the error polar diagram. The constant of proportionality depends upon the shape of ϕ_q and well known formulas[8] can be used for determining it. Thus, the magnitude and extent of spurious side lobes due to the errors can be estimated. Since the actual autocorrelation function of the errors has been calculated the angular direction of the side lobes due to them can be predicted.

Conclusions

Two approximate but general formulas, (16) and (17), have been given for the lowest probable side-lobe level and reduction of gain for antennas exhibiting random manufacturing errors. The effect of the errors on side-lobe level is much the more important.

The theory given here supports the view that the way in which an aperture is illuminated (the design of the primary feed for a reflector, for instance) is relatively less important than the accuracy with which the antenna is made.

It is worth noting that $|S(\psi)|$, in a two-dimensional analysis of the aperture such as Ruze's,[4] would depend upon the square of the radius of correlation rather than upon the first power as here. The reason for the difference is that in this note the aperture illumination has been collapsed into an equivalent line source in (1) and (7). This is justified because in practice it is only possible to measure polar diagrams in one plane at a time, and in the far field a two-dimensional antenna and its equivalent line source (lying in the plane in which the polar diagram is being measured) appear to be the same.

Acknowledgment

The permission of Canadian Westinghouse Co., Ltd. to publish this note is gratefully acknowledged.

[8] S. Silver, *op. cit.*, p. 187.

Successive Variational Approximations of Impedance Parameters in a Coupled Antenna System*

M. K. HU† AND Y. Y. HU†

Summary—**In this paper, a new variational formulation for a single impedance parameter of an m-antenna system is presented. This formulation enables one to determine any self impedance Z_{ii}, one at a time, merely by exciting antenna i alone and leaving all the other antennas open circuited. For determining any mutual impedance Z_{ij}, only two independent excitations, one the same as that used for determining Z_{ii} and the other for determining Z_{jj}, are required. Thus, if all the $m(m+1)/2$ impedance are required, only m independent excitation conditions are needed. In contrast to this, the formulation available in the literature is based on $m(m+1)/2$ independent excitation conditions. Because of a reduced number of excitation conditions and the way they are assumed, the physical nature of the problem is made simpler and easier to comprehend. Such comprehension helps considerably in the choice of trial current distributions for a specific application.**

Two methods of evaluating the successive higher-order approximations are also given. One is based upon an orthogonalization process, and the other is based upon the successive inversion of matrices. In the evaluation of a certain order approximation, both methods have the advantage of utilizing all the work already done for the lower-order approximations; and at the same time, additional work required is considerably reduced. It is believed that the formulation, as well as the two methods of successive approximations, will also be useful in other problems.

Introduction

FOR A SYSTEM of coupled antennas, the terminal electrical behavior can be described by the self and mutual impedances of the coupled antennas in the system. A comprehensive review of different methods for determining the impedance parameters of antenna arrays was given by Levis.[1] Levis and Tai also published a variational formulation for obtaining the $m(m+1)/2$ impedance parameters of an m-antenna system by solving a set of simultaneous equations derived from $m(m+1)/2$ "independent" excitation conditions.[2] In the present paper, a formulation is presented; it gives a general variational expression for a single impedance parameter, which can be any one of the $m(m+1)/2$ parameters. This formulation is considerably simpler and straightforward; as a consequence, it gives a deeper insight to the physical nature of the problem. It increases greatly the flexibility in the choice of the trial current distributions for a particular impedance parameter. If all the impedance parameters are required, it involves only m independent excitation conditions. For a particular mutual impedance, only two independent conditions are required; and for a particular self-impedance, only one condition is necessary. Formulas which are expressed in terms of trial current distributions and the geometry of the system are also derived by eliminating the adjustable coefficients according to the stationary property of the general variational expression. These formulas give the impedance parameter directly, instead of through solving the admittance parameters as an intermediate step; therefore, the computations involved are very much reduced in actual applications.

It is also shown that the above impedance formula can be further converted into an orthogonalized form by a modified Schmidt process or the "bi-orthogonalization" process. This orthogonalized form is very useful in evaluating successive higher-order approximations of the impedance parameter. It has the advantage of utilizing all the work already done for the lower-order approximations; and, at the same time, it reduces the additional work required for the next higher-order approximation. In this formula, only one additional term is involved for the next higher-order approximation. If the terms are evaluated successively, the rate of convergence of the approximations can be seen clearly and used as a guide for the choice of the higher-order trial functions. This formula is also useful for the estimation of the accuracy of a particular variational solution. In such applications, additional trial function can be assumed, from either physical or mathematical considerations; then the order of magnitude of the additional term is an estimation of the error involved in the original approximation.

Another method of evaluating successive higher-order approximations, which is also given, is based upon inversion of matrices of successive higher orders. Basically, this method is equivalent to the method of orthogonalization; the implications are about the same, but each does possess certain merits over the other in actual applications.

In the following discussion, parallel cylindrical antennas will be used for the convenience of presentation, but the method can be easily extended to arbitrary orientation and also to other types of antennas. Furthermore, the bi-orthogonalization process and the method of successive inversion can be applied to a variety of problems wherever similar mathematical form occurs.

Variational Formulation and Variational Expression

The present formulation is to obtain a variational expression for any single impedance parameter of an m-

* Manuscript received by the PGAP, July 28, 1958; revised manuscript received, March 20, 1959.

† Elec. Engrg. Dept., Syracuse University, Syracuse, N. Y.

[1] C. A. Levis, "Variational Calculations of the Impedance Parameters of Coupled Antennas," Antenna Lab., Ohio State Univ., Columbus, Rept. No. 667-16; August, 1956.

[2] C. A. Levis and C. T. Tai, "A method of analyzing coupled antennas of unequal sizes," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-4, pp. 128–132; April, 1956.

antenna system, antennas i and j are chosen arbitrarily for the discussion of the mutual impedance Z_{ij}. The m linear antennas are all assumed to be oriented in the same direction, say z-direction, as shown by Fig. 1. For

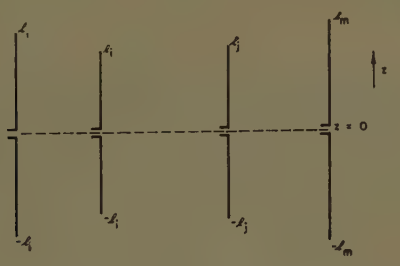


Fig. 1—A system of m coupled linear antennas.

simplicity in discussion, it is also assumed that all the antennas are center-fed and all the feeding points are located in the $z=0$ plane. When antenna i is connected to a current source of unit amplitude and with all the other antennas open-circuited, the current distributions on the antennas 1 to m will be denoted by $I_1^i(z)$, $I_2^i(z)$, $\cdots$, $I_m^i(z)$. Similarly, when antenna j is connected to a current source of unit amplitude and with all the other antennas open-circuited, the current distributions on the antennas 1 to m will be denoted by $I_1^j(z)$, $I_2^j(z)$, $\cdots$, $I_m^j(z)$. The above implies the following relations:

$$\begin{aligned} I_p^i(0) &= 0 \qquad \text{when } p \neq i \\ I_p^i(0) &= 1 \qquad \text{when } p = i \end{aligned} \tag{1}$$

and

$$\begin{aligned} I_q^j(0) &= 0 \qquad \text{when } q \neq j \\ I_q^j(0) &= 1 \qquad \text{when } q = j. \end{aligned} \tag{2}$$

These relations may be called the feeding conditions. In addition, the current distributions are further assumed to satisfy the following end conditions as generally used in linear antenna theory:

$$I_p^i(\pm l_p) = 0$$

and

$$I_q^i(\pm l_q) = 0 \qquad p, q = 1, 2, \cdots, m \tag{3}$$

where l_p and l_q are the half-lengths of antennas p and q, respectively.

In linear cylindrical antenna theory, it is also known that the set of current distributions due to $I_i^i(0)$ alone satisfies the following integral equations:

$$V_q^i\delta(z) = \sum_{p=1}^{m} \int_{-l_p}^{l_p} G(z, z') I_p^i(z') dz', \quad q = 1, 2, \cdots, m \tag{4}$$

where $\delta(z)$ is the Dirac delta function, V_q^i is the voltage appearing across the feeding terminals of antenna q due to feeding current $I_i^i(0)$ alone at antenna i, $G(z, z')$ is the free-space Green's function given by

$$G(z, z') = \frac{j\beta\eta}{4\pi}\left(1 + \frac{1}{\beta^2}\frac{\partial^2}{\partial z^2}\right)\frac{e^{-i\beta r}}{r}$$

with

$$\beta = \frac{2\pi}{\lambda}, \qquad \eta = \sqrt{\mu_0/\epsilon_0},$$

and r = the distance between z and z' on antennas q and p respectively. Similarly, the set of current distributions due to feeding current $I_j^j(0)$ alone satisfies a similar set of integral equations:

$$V_p^j\delta(z) = \sum_{q=1}^{m} \int_{-l_q}^{l_q} G(z, z') I_q^j(z') dz', \quad p = 1, 2, \cdots, m. \tag{5}$$

According to the usual definitions of antenna impedances, V_q^i and V_p^j in (4) and (5) are related to the mutual impedances Z_{qi} and Z_{pj} by

$$V_q^i = Z_{qi} I_i^i(0), \qquad q = 1, 2, \cdots, m. \tag{6}$$

$$V_p^j = Z_{pj} I_j^j(0), \qquad p = 1, 2, \cdots, m. \tag{7}$$

Multiplying (4) by $I_q^j(z)$ and integrating from $-l_q$ to l_q, and further using (6), we have

$$I_q^j(0) Z_{qi} I_i^i(0) = \sum_{p=1}^{m} \int_{-l_q}^{l_q} \int_{-l_p}^{l_p} I_q^j(z) G(z, z') I_p^i(z') dz' dz$$
$$q = 1, 2, \cdots, m. \tag{8}$$

Similarly, from (5) and (7), we have

$$I_p^i(0) Z_{pj} I_j^j(0) = \sum_{q=1}^{m} \int_{-l_p}^{l_p} \int_{-l_q}^{l_q} I_p^i(z) G(z, z') I_q^j(z') dz' dz$$
$$p = 1, 2, \cdots, m. \tag{9}$$

Summing (8) for $q=1, 2, \cdots, m$, and noting that $I_q^j(0)=1$ for $q=j$ and $I_q^j(0)=0$ for $q\neq j$, we obtain

$$Z_{ji} = \sum_{q=1}^{m} \sum_{p=1}^{m} \int_{-l_q}^{l_q} \int_{-l_p}^{l_p} I_q^j(z) G(z, z') I_p^i(z') dz' dz. \tag{10}$$

A similar equation can be obtained from (9),

$$Z_{ij} = \sum_{p=1}^{m} \sum_{q=1}^{m} \int_{-l_p}^{l_p} \int_{-l_q}^{l_q} I_p^i(z) G(z, z') I_q^j(z') dz' dz. \tag{11}$$

It can be shown that both Z_{ji} and Z_{ij} given by (10) and (11) are variational forms. In fact, $Z_{ji}=Z_{ij}$, since $G(z, z')=G(z', z)$. Therefore, (10) and (11) are actually equivalent to each other; so in the following discussion, only (11) will be used.

It is obvious that if the two sets of current distributions $I_1^i(z)$, $I_2^i(z)$, $\cdots$, $I_m^i(z)$ and $I_1^j(z)$, $I_2^j(z)$, $\cdots$, $I_m^j(z)$ are known exactly, then there is no particular advantage of using (11) for the determination of Z_{ij} over other simpler methods. The real advantage of using (11) is in those cases where the current distributions are unknown or only known approximately. Because Z_{ij} given by (11) is a variational form, Z_{ij} is stationary with respect to the true current distributions. Literally speak-

ing, this means that if the approximate current distributions have an error of certain order, the impedance computed by using (11) will generally have an error of higher order—a smaller error.

The stationary property of Z_{ij} can be proved by taking the first variation of (11),

$$\delta Z_{ij} + \delta I_i^i(0) Z_{ij} + Z_{ij}\delta I_j^j(0) = \sum_{p=1}^{m}\sum_{q=1}^{m}\int_{-l_p}^{l_p}\int_{-l_q}^{l_q}\delta I_p^i(z)G(z, z')I_q^j(z')dz'dz + \sum_{p=1}^{m}\sum_{q=1}^{m}\int_{-l_p}^{l_p}\int_{-l_q}^{l_q} I_p^i(z)G(z, z')\delta I_q^j(z')dz'dz. \quad (12)$$

By using (5) and (7), and the relation $\delta I_p^i(0)=0$ for $p\neq i$, the first double sum of (12) gives

$$\sum_{p=1}^{m}\sum_{q=1}^{m}\int_{-l_p}^{l_p}\int_{-l_q}^{l_q}\delta I_p^i(z)G(z, z')I_q^j(z')dz'dz = \delta I_i^i(0)Z_{ij}. \quad (13)$$

Similarly, by using (4) and (6), the second double sum of (12) gives,

$$\sum_{p=1}^{m}\sum_{q=1}^{m}\int_{-l_p}^{l_p}\int_{-l_q}^{l_q} I_p^i(z)G(z, z')\delta I_q^j(z')dz'dz = Z_{ij}\delta I_j^j(0). \quad (14)$$

Subtracting (13) and (14) from (12), we obtain

$$\delta Z_{ij} = 0.$$

This means that there is no first variation of Z_{ij} given by (11) because of the first variations of current distributions. The stationary property of Z_{ij} is therefore proved. Since i and j are arbitrarily chosen, therefore (11) is a general variational expression for mutual impedance ($i\neq j$) or self-impedance ($i=j$) of the coupled antennas.

In the above proof, both $\delta I_i^i(0)$ and $\delta I_j^j(0)$ are allowed to take non-zero values as well as zero values. Physically, the consideration of non-zero variations $\delta I_i^i(0)$ and $\delta I_j^j(0)$ can also be treated as a change in the excitation levels from $I_i^i(0)$ and $I_j^j(0)$ to $I_i^i(0)+\delta I_i^i(0)$ and $I_j^j(0)+\delta I_j^j(0)$. In a linear system, the governing equations and parameters are independent of the excitation levels; therefore, there is no loss of generality in assuming that both $\delta I_i^i(0)$ and $\delta I_j^j(0)$ are zero when the variations of the current distributions are considered. Under this assumption, the manipulations involved in actual applications and discussions to be presented in the next section are considerably simplified. In fact, if special trial functions are assumed, as to be described below, it can be shown that the result obtained by using this assumption is identically equal to the result obtained without using it.

Formulas for Z_{ij} in Terms of Trial Current Distributions

In order to take full advantage of the stationary property of (11), adjustable constants are usually incorporated in the assumed approximate current distributions. These constants are then determined by using the condition that the first variation of the impedance parameter is zero. In practice, these adjustable constants appear only linearly in the assumed current distributions; in other words, the assumed current distribution on each antenna appears as a linear combination of several trial functions with adjustable coefficients.

For the set of current distributions $I_1^i(z), I_2^i(z), \cdots, I_m^i(z)$, used in the last section, each distribution including $I_i^i(z)$ on antenna i may contain several trial functions and as many adjustable coefficients. For the purpose of simplifying the notation, all these trial functions except one will be arranged in an arbitrary order and denoted by $I_1(z), I_2(z), \cdots, I_m(z), \cdots, I_n(z)$, with the corresponding adjustable coefficients denoted by $k_1^i, k_2^i, \cdots, k_m^i, \cdots, k_n^i$. According to the feeding and end conditions given in the last section, the following conditions will be imposed on these n trial functions:

$$I_\mu(0) = I_\mu(\pm l_\mu) = 0, \qquad \mu = 1, 2, \cdots, n \quad (15)$$

where l_μ denotes the half-length of the particular antenna on which the current $I_\mu(z)$ is located. If $I_\mu(z)$ is located on antenna p, then $l_\mu = l_p$.

In addition to these n trial functions, one more trial function, $I_0(z)$, will be assumed for $I_i^i(z)$ to take care of the feeding condition on antenna i. $I_0(z)$ is assumed to satisfy the following conditions:

$$I_0(0) = 1$$

and

$$I_0(\pm l_i) = 0. \quad (16)$$

In general, one more adjustable coefficient, k_0^i, may be assumed for $I_0(z)$, but it is not necessary. As discussed in the last section, there is no loss of generality, if $\delta I_i^i(0)$ is taken to be zero. This implies that k_0^i can be considered as a constant. Furthermore, because of the unity feeding condition, k_0^i can then be taken as 1. It should be noted that the choice of $I_0(z)$ is just as flexible as any of the other trial functions except that it is required to satisfy $I_0(0)=1$ instead of zero.

Similarly, for the other set of current distributions, $I_1^j(z), \cdots, I_m^j(z)$, the following set of trial functions and the corresponding coefficients will be assumed and denoted by

$$J_0(z), J_1(z), \cdots, J_m(z), \cdots, J_{n'}(z)$$

and

$$k_0^j = 1, k_1^j, \cdots, k_m^j, \cdots, k_{n'}^j.$$

This set of trial functions is assumed to satisfy a similar set of feeding and end conditions as given by (15) and (16). Some of these functions could be the same as some of those assumed in the first set, or they could be entirely different. Generally, n' may not be equal to n, but in actual application, it is preferred to have $n'=n$.

Using these two sets of trial current distributions, and

denoting the approximate solution of Z_{ij} by $\hat{Z}_{ij}$, (11) then becomes:

$$\hat{Z}_{ij} = \sum_{\mu=0}^{n} \sum_{\nu=0}^{n'} \int_{-l_\mu}^{l_\mu} \int_{-l_\nu}^{l_\nu} k_\mu{}^i I_\mu(z) G(z, z') k_\nu{}^j J_\nu(z') dz' dz \quad (17)$$

where l_μ and l_ν are the half-lengths of the particular antennas on which the currents $I_\mu(z)$ and $I_\nu(z)$ are located.

In the following discussion, matrix notation will be used. It is simpler and the relations expressed in matrix form are much easier to comprehend. Eq. (17) in matrix form reads:

$$\hat{Z}_{ij} = [k^i][(I, J)][k^j] \quad (18)$$

where

$$[k^i] = [1, k_1{}^i, \cdots, k_n{}^i]$$

$$[k^j] = \begin{bmatrix} 1 \\ k_1{}^j \\ \cdots \\ k_{n'}{}^j \end{bmatrix}$$

and

$$[(I, J)] = \begin{bmatrix} (I_0, J_0) & (I_0, J_1) \cdots (I_0, J_{n'}) \\ (I_1, J_0) & (I_1, J_1) \cdots (I_1, J_{n'}) \\ \cdots & \cdots \\ (I_n, J_0) & (I_n, J_1) \cdots (I_n, J_{n'}) \end{bmatrix},$$

with

$$(I_\mu, J_\nu) = \int_{-l_\mu}^{l_\mu} \int_{-l_\nu}^{l_\nu} I_\mu(z) G(z, z') J_\nu(z') dz' dz.$$

The adjustable coefficients k's can then be determined by taking the first variation of (18). This is

$$\delta\hat{Z}_{ij} = [\delta k^i][(I, J)][k^j] + [k^i][(I, J)][\delta k^j] \quad (19)$$

where

$$[\delta k^i] = [0, \delta k_1{}^i, \cdots, \delta k_n{}^i]$$

and

$$[\delta k^j] = \begin{bmatrix} 0 \\ \delta k_1{}^j \\ \cdots \\ \delta k_{n'}{}^j \end{bmatrix}.$$

In order to have $\delta\hat{Z}_{ij}=0$, it is required that all the coefficients of the variation δk's should be set equal to zero, therefore the following two sets of equations are obtained

$$\begin{bmatrix} (I_1, J_0) & (I_1, J_1) \cdots (I_1, J_{n'}) \\ \cdots & \cdots \\ (I_n, J_0) & (I_n, J_1) \cdots (I_n, J_{n'}) \end{bmatrix} \begin{bmatrix} 1 \\ k_1{}^j \\ \cdots \\ k_{n'}{}^j \end{bmatrix} = \begin{bmatrix} 0 \\ \cdots \\ 0 \end{bmatrix} \quad (20)$$

and

$$[1, k_1{}^i, \cdots, k_n{}^i] \begin{bmatrix} (I_0, J_1) \cdots (I_0, J_{n'}) \\ (I_1, J_1) \cdots (I_1, J_{n'}) \\ \cdots \\ (I_n, J_1) \cdots (I_n, J_{n'}) \end{bmatrix} = [0, \cdots, 0]. \quad (21)$$

Eqs. (20) and (21) can also be written as:

$$\begin{bmatrix} (I_1, J_1) \cdots (I_1, J_{n'}) \\ \cdots \\ (I_n, J_1) \cdots (I_n, J_{n'}) \end{bmatrix} \begin{bmatrix} k_1{}^j \\ \cdots \\ k_{n'}{}^j \end{bmatrix} = - \begin{bmatrix} (I_1, J_0) \\ \cdots \\ (I_n, J_0) \end{bmatrix} \quad (22)$$

and

$$[k_1{}^i, \cdots, k_n{}^i] \begin{bmatrix} (I_1, J_1) \cdots (I_1, J_{n'}) \\ \cdots \\ (I_n, J_1) \cdots (I_n, J_{n'}) \end{bmatrix} = - [(I_0, J_1) \cdots (I_0, J_n)]. \quad (23)$$

It is well known that in the theory of simultaneous equations, the following two conditions:

$$n' = n$$

and

$$\begin{vmatrix} (I_1, J_1) \cdots (I_1, J_n) \\ \cdots \\ (I_n, J_1) \cdots (I_n, J_n) \end{vmatrix} \neq 0 \quad (24)$$

imply the existence of unique solutions for k^i, k^j in (22) and (23). If the solutions are worked out and substituted back into (18), we have

$$\hat{Z}_{ij} = (I_0, J_0) - [(I_0, J_1), \cdots, (I_0, J_n)] \cdot \begin{bmatrix} (I_1, J_1) \cdots (I_1, J_n) \\ \cdots \\ (I_n, J_1) \cdots (I_n, J_n) \end{bmatrix}^{-1} \begin{bmatrix} (I_1, J_0) \\ \cdots \\ (I_n, J_0) \end{bmatrix}. \quad (25)$$

Another formula, which is equivalent to (25), can also be derived:

$$\hat{Z}_{ij} = \frac{\begin{vmatrix} (I_0, J_0) & (I_0, J_1) \cdots (I_0, J_n) \\ (I_1, J_0) & (I_1, J_1) \cdots (I_1, J_n) \\ \cdots & \cdots \\ (I_n, J_0) & (I_n, J_1) \cdots (I_n, J_n) \end{vmatrix}}{\begin{vmatrix} (I_1, J_1) \cdots (I_1, J_n) \\ \cdots \\ (I_n, J_1) \cdots (I_n, J_n) \end{vmatrix}}. \quad (26)$$

Both formulas (25) and (26) give the mutual impedance directly instead of indirectly through the solutions of admittances. These formulas are expressed in terms of the trial current distributions and the geometry of the system only. They can be applied to the mutual impedances ($i \neq j$) as well as the self-impedances ($i = j$).

In the self-impedance case, it is clear that all J's may be replaced by I's.

In actual applications, the two conditions given by (24) can be easily satisfied and should be used. In the seemingly more general cases where either $n' \neq n$ or the determinant is equal to zero, there are only additional complications but with no contribution to any improvement of the solution. This will be discussed in the Appendix.

In the computation of the impedances $\hat{Z}_{ii}$, $\hat{Z}_{jj}$ and $\hat{Z}_{ij}$, if the two sets of trial functions are chosen such that,

$$I_p(z) = J_p(z) \qquad \text{for } p = 1, \cdots, n$$

and with the corresponding currents $I_p(z)$ and $J_p(z)$ located on the same antenna, then the current distributions determined from the stationary property of $\hat{Z}_{ij}$ will be identical to those determined from the stationary properties of $\hat{Z}_{ii}$ and $\hat{Z}_{jj}$. This can be seen clearly from (22) and (23).

The Orthogonalized Formula for $\hat{Z}_{ij}$

The formula for $\hat{Z}_{ij}$ given by (25) can be further reduced to an orthogonalized form, if the two sets of trial current functions are orthogonalized by a modified Schmidt process. This modified process will be called the "bi-orthogonalization" process, and can be expressed systematically as follows:

$$\begin{aligned}
x_1 &= I_1 \\
x_2 &= I_2 - \frac{(I_2, y_1)}{(x_1, y_1)} x_1 \\
x_3 &= I_3 - \frac{(I_3, y_1)}{(x_1, y_1)} x_1 - \frac{(I_3, y_2)}{(x_2, y_2)} x_2 \\
&\cdots\cdots\cdots\cdots\cdots \\
x_n &= I_n - \frac{(I_n, y_1)}{(x_1, y_1)} x_1 - \frac{(I_n, y_2)}{(x_2, y_2)} x_2 - \cdots \\
&\quad - \frac{(I_n, y_{n-1})}{(x_{n-1}, y_{n-1})} x_{n-1}
\end{aligned}$$

and

$$\begin{aligned}
y_1 &= J_1 \\
y_2 &= J_2 - \frac{(x_1, J_2)}{(x_1, y_1)} y_1 \\
y_3 &= J_3 - \frac{(x_1, J_3)}{(x_1, y_1)} y_1 - \frac{(x_2, J_3)}{(x_2, y_2)} y_2 \\
&\cdots\cdots\cdots\cdots\cdots \\
y_n &= J_n - \frac{(x_1, J_n)}{(x_1, y_1)} y_1 - \frac{(x_2, J_n)}{(x_2, y_2)} y_2 - \cdots \\
&\quad - \frac{(x_{n-1}, J_n)}{(x_{n-1}, y_{n-1})} y_{n-1}.
\end{aligned} \tag{27}$$

It should be noted that the operations associated with the subtraction signs appearing in the above equations should not be carried out at this stage, because the same notation z is used for the different variables associated with different antennas. These operations should be carried out only after the double integrations are evaluated with the locations of the currents.

From (27), it is easy to prove that the following relations are true

$$(x_\mu, y_\nu) = 0 \qquad \text{if } \mu \neq \nu$$
$$\mu, \nu = 1, 2, \cdots, n.$$

In the above process, all the quantities (x_1, y_1), (x_2, y_2), $\cdots$, (x_n, y_n) are assumed to be unequal to zero. If at a certain stage, this does not occur, say $(x_\mu, y_\mu) = 0$, then the situation is similar to that discussed in the Appendix. It can be proved that the solution will not be changed by the particular pair of trial functions $I_\mu(z)$ and $J_\mu(z)$, so they can be omitted and a new pair of trial functions be assumed such that $(x_\mu, y_\mu) \neq 0$.

If the two sets of new functions $x_1, x_2, \cdots, x_n$ and $y_1, y_2, \cdots, y_n$ are used as the new trial functions, and $1, \bar{k}_1^i, \cdots, \bar{k}_n^i$ and $1, \bar{k}_1^j, \cdots, \bar{k}_n^j$ as the new adjustable coefficients, then (18) can be written as:

$$\hat{Z}_{ij} = [\bar{k}^i][(x, y)][\bar{k}^j], \tag{28}$$

where

$$[(x, y)] = \begin{bmatrix} (x_0, y_0) & (x_0, y_1) & (x_0, y_2) & \cdots & (x_0, y_n) \\ (x_1, y_0) & (x_1, y_1) & 0 & \cdots & 0 \\ (x_2, y_0) & 0 & (x_2, y_2) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ (x_n, y_0) & 0 & 0 & \cdots & (x_n, y_n) \end{bmatrix}$$

with

$$x_0 = I_0, \qquad y_0 = J_0.$$

If the same method of evaluating $[k^i]$ and $[k^j]$ is used here, then we have the following solution:

$$\hat{Z}_{ij} = (x_0, y_0) - [(x_0, y_1) \cdots (x_0, y_n)] \cdot \begin{bmatrix} (x_1, y_1) & \cdots & 0 \\ \cdots & \cdots & \cdots \\ 0 & \cdots & (x_n, y_n) \end{bmatrix}^{-1} \begin{bmatrix} (x_1, y_0) \\ \cdots \\ (x_n, y_0) \end{bmatrix}.$$

This can also be written simply as

$$\hat{Z}_{ij} = (x_0, y_0) - \frac{(x_0, y_1)(x_1, y_0)}{(x_1, y_1)} - \cdots - \frac{(x_0, y_n)(x_n, y_0)}{(x_n, y_n)}. \tag{29}$$

The solutions for $\hat{Z}_{ij}$ given by (25) and (29) are equal to each other. This can be proved directly by mathematical induction or by applying the following theorem.

Theorem:[3] If two different systems of trial functions

[3] The proofs of this theorem and some more general theorems are being prepared for publication.

are related by non-singular linear transformations, then the two variational solutions obtained for $\hat{Z}_{ij}$ are equal.

The advantage of (29) over (25) is fairly obvious. For the evaluation of the next higher-order approximation, only one additional term is involved. It does not need the inversion of matrix as required by (25). In addition, it has the advantage of utilizing all the work already done for the lower-order approximations. As a result, it reduces considerably the work required in evaluating higher-order approximations. Moreover, the terms in (29) show clearly the rate of convergence of the different order of approximations. This formula is also useful for the estimation of the accuracy of a particular variational solution. In such applications, additional trial function can be assumed from either physical or mathematical considerations; then the order of magnitude of the additional term is an estimation of the error involved.

Evaluation of $\hat{Z}_{ij}$ by Successive Inversion

The approximation given previously by (25) may be called the nth order approximation of Z_{ij}. For the convenience of the following discussion, it will be given here again and denoted by $Z_{ij}^{(n)}$:

$$Z_{ij}^{(n)} = (I_0, J_0) - [(I_0, J_1) \cdots (I_0, J_n)] \cdot \begin{bmatrix} (I_1, J_1) & \cdots & (I_1, J_n) \\ \cdot & \cdots & \cdot \\ (I_n, J_1) & \cdots & (I_n, J_n) \end{bmatrix}^{-1} \begin{bmatrix} (I_1, J_0) \\ \cdots \\ (I_n, J_0) \end{bmatrix}. \quad (30)$$

It is obvious that the above form applies for any n. In finding $Z_{ij}^{(n)}$ for a particular problem, besides the evaluation of the double integrals, it is necessary to invert an nth order matrix. The inversion of a matrix is a time consuming process; it is especially true when n is large. The method to be presented here reduces the required work considerably, being based upon a successive-inversion procedure. This procedure can be applied for the inversion of any matrix, but it is particularly advantageous in the present or similar applications. It not only gives the nth order approximation, but also provides all the lower-order approximations in the process of evaluation. The rate of convergence of these lower-order approximations can be seen clearly just as in using the method given in the previous section. If still higher-order approximations are required, the same procedure can be applied without additional difficulty.

This successive-inversion procedure is based upon the following theorem.

Theorem: If a given nth order matrix $A^{(n)}$ is partitioned and written in compound matrix form as:

$$A^{(n)} = \begin{bmatrix} A_{\alpha\alpha} & A_{\alpha n} \\ A_{n\alpha} & A_{nn} \end{bmatrix}$$

and its $(n-1)$th order submatrix $A_{\alpha\alpha}$ is assumed to have a known inverse, denoted by $B_{\alpha\alpha}^{(n-1)}$; then the inverse of $A^{(n)}$, denoted by $B^{(n)}$, in compound matrix form is given by:

$$B^{(n)} = \begin{bmatrix} B_{\alpha\alpha}^{(n)} & B_{\alpha n}^{(n)} \\ B_{n\alpha}^{(n)} & B_{nn}^{(n)} \end{bmatrix} \quad (31)$$

with

$$\begin{aligned} B_{nn}^{(n)} &= (A_{nn} - A_{n\alpha} B_{\alpha\alpha}^{(n-1)} A_{\alpha n})^{-1} \\ B_{n\alpha}^{(n)} &= - B_{nn}^{(n)} A_{n\alpha} B_{\alpha\alpha}^{(n-1)} \\ B_{\alpha n}^{(n)} &= - B_{\alpha\alpha}^{(n-1)} A_{\alpha n} B_{nn}^{(n)} \\ B_{\alpha\alpha}^{(n)} &= B_{\alpha\alpha}^{(n-1)} + B_{\alpha\alpha}^{(n-1)} A_{\alpha n} B_{nn}^{(n)} A_{n\alpha} B_{\alpha\alpha}^{(n-1)}. \end{aligned}$$

The proof of the above theorem is straightforward and will be omitted here. The application of this theorem to the evaluation of (30) is obvious; n can be set first equal to 2, then to 3, and so on.

The present method is basically equivalent to the one given in the previous section, but it is in a more organized and easily applicable form, because of the use of matrix operations. On the other hand, the contribution of each trial function can be seen better in the method of the previous section through (27); this fact is helpful in the choice of the next higher order trial function.

Appendix

On Solutions of $\hat{Z}_{ij}$

The two conditions given by (24) are sufficient for the existence of consistent and unique solutions for k^i and k^j in (22) and (23). But from the theory of linear simultaneous equation, a necessary and sufficient condition for the existence of consistent solutions is that the following three matrices

$$\begin{bmatrix} (I_1, J_1) & \cdots & (I_1, J_{n'}) \\ \cdot & \cdots & \cdot \\ (I_n, J_1) & \cdots & (I_n, J_{n'}) \end{bmatrix}, \quad \begin{bmatrix} (I_1, J_0) & (I_1, J_1) & \cdots & (I_1, J_{n'}) \\ \cdot & \cdot & \cdots & \cdot \\ (I_n, J_0) & (I_n, J_1) & \cdots & (I_n, J_{n'}) \end{bmatrix},$$

$$\begin{bmatrix} (I_0, J_1) & \cdots & (I_0, J_{n'}) \\ (I_1, J_1) & \cdots & (I_1, J_{n'}) \\ \cdot & \cdots & \cdot \\ (I_n, J_1) & \cdots & (I_n, J_{n'}) \end{bmatrix}$$

are all of the same rank, say r.

From this condition, it is then always possible to find at least one non-zero determinant of order r from the first matrix. By rearranging and renumbering the rows and columns of the first matrix, it is always possible to have

$$\begin{bmatrix} (I_1, J_1) & \cdots & (I_1, J_r) \\ \cdot & \cdots & \cdot \\ (I_r, J_1) & \cdots & (I_r, J_r) \end{bmatrix} \neq 0.$$

Eqs. (22) and (23) should be rearranged and renumbered accordingly. After this is done, it is possible to solve the first r coefficients $k_1^i, \cdots, k_r^i$ in terms of the remaining $(n'-r)$ coefficients $k_{r+1}^i, \cdots, k_{n'}^i$ in (22); and similarly, the first r coefficient $k_1^i, \cdots, k_r^i$ in terms of the remaining $(n-r)$ coefficients $k_{r+1}^i, \cdots, k_n^i$ in (23). If these are worked out and substituted back into (18), after simplifying, we then have

$$\hat{Z}_{ij} = (I_0, J_0) - [(I_0, J_1) \cdots (I_0, J_r)] \cdot \begin{bmatrix} (I_1, J_1) \cdots (I_1, J_r) \\ \cdots \\ (I_r, J_1) \cdots (I_r, J_r) \end{bmatrix}^{-1} \begin{bmatrix} (I_1, J_0) \\ \cdots \\ (I_r, J_0) \end{bmatrix}. \quad (32)$$

It should be noted that (32) is finally independent on the remaining $(n+n'-2r)$ coefficients $k_{r+1}^i, \cdots, k_n^i$ and $k_{r+1}^j, \cdots, k_{n'}^j$, and the corresponding trial functions. In actual application, r is generally equal to the smaller one of the two numbers, n and n'. If n is the smaller one of the two, then (32) is identical to (25). But it should also be noted that the additonal $(n'-n)$ trial functions cannot be chosen arbitrarily, because they must satisfy the above necessary and sufficient condition; otherwise there will be no consistent solutions at all for (22) and (23). Furthermore, the condition is a very stringent one. Even if it is satisfied, it does not contribute to any improvement of the solution. Therefore, it is clear that n' should be chosen equal to n in actual applications.

A New Method for Obtaining Maximum Gain from Yagi Antennas*

H. W. EHRENSPECK† AND H. POEHLER†

Summary—**In conventional Yagi design, optimum performance requires separate adjustments in a number of parameters—the array length and the height, diameter, and spacing of the directors and reflectors.**

By introducing the notion of a surface wave traveling along the array, it is possible to demonstrate experimentally the interrelationship between these parameters. With this, the gain then depends only on the phase velocity of the surface wave (which is a function of the height, diameter, and spacing of the directors) and on the choice of the reflector. Thus, maximum gain for a given array length, for any director spacing less than 0.5 λ, can be obtained by suitable variation of the parameters to yield the desired phase velocity.

A design procedure that provides maximum gain for a given array length is presented.

Introduction

A Yagi antenna consists of a number of linear elements parallel to each other along a straight line. Fig. 1 represents (a) a conventional Yagi antenna, and (b) one that consists of a row of monopoles imaged in a ground plane. Only one of the elements—the so-called *feeder*—is excited, all the others being parasitic. The elements in the direction of increased gain are directors, and the one in the opposite direction is the reflector. By choosing the correct dimensions for the different parameters of the Yagi antenna (height, spacing, and diameter of the elements), we can obtain increased gain in the endfire direction of the array.

The simplicity of Yagi antenna construction has invited many applications. One of the advantages of the Yagi is its usefulness as practically all frequency bands.

In designing a Yagi antenna the problem is to find the correct dimensions of the parameters. It is therefore necessary to discover how these parameters are related to each other under conditions of optimum gain.

Some early investigators of this problem, Yagi [1], Uda and Mushiake [2], and Walkinshaw [3], computed the currents in the elements according to the impedance concept but treated only short Yagi antennas because the computation for more than three or four elements becomes too complicated. Others, Reid [4], Fishenden and Wiblin [5], and Vysokovsky [6] [7], incorrectly, as first pointed out by Vysokovsky [6], [7], assumed that the current amplitudes were the same in all the directors. To date, there exists no rigorous theoretical solution of the Yagi problem. The experimental results that have been published are always restricted to special cases, with no attempt made to find a connection between them.

The goal of our investigation was to establish a general design method for Yagi antennas with optimum gain. The investigation was restricted to a consideration of Yagis with a single reflector and directors of equal height, spacing, and diameter. Our design criteria were based on achieving forward gain without respect to the front-to-back ratio and the sidelobe level.

* Manuscript received by the PGAP, January 30, 1958; revised manuscript received, February 20, 1959.

† Electromagnetic Radiation Lab., Electronics Res. Directorate, Air Force Cambridge Res. Center, Air Res. and Dev. Command, Bedford, Mass.

All gain figures are given in terms of power gain above a reference dipole (the feeder) and they thus include the effect of ohmic losses. The symbols used are (see Fig. 2):

h_F = height of feeder above ground plane
$2\rho_F$ = diameter of feeder
h_R = height of reflector above ground plane
$2\rho_R$ = diameter of reflector
s_R = spacing between reflector and feeder
h_D = height of directors above ground plane
$2\rho_D$ = diameter of directors
s_D = spacing of directors
L = length of the array (measured from feeder to last director)
n = number of directors
λ = wavelength
v_x = phase velocity along the antenna
c = phase velocity in free space
g = power gain over dipole
g_{max} = maximum power gain of Yagi antenna.

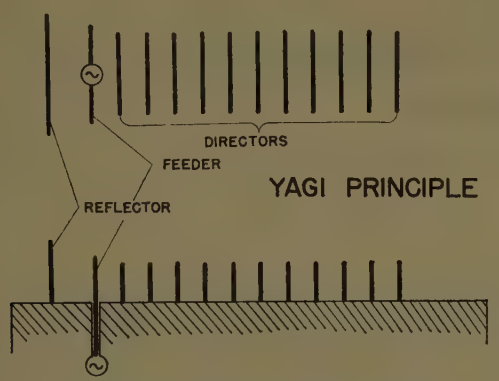


Fig. 1—(a) Conventional Yagi; (b) Yagi consisting of monopoles imaged in a ground plane.

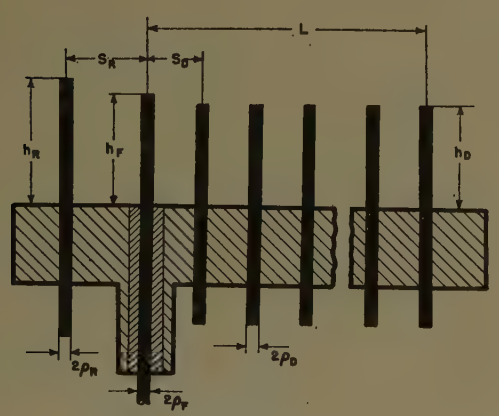


Fig. 2—Yagi parameter symbols.

Treatment of the Problem

Our basic approach was to consider the Yagi antenna as consisting of two parts: the combination feeder-reflector and the row of director elements. Experimental evidence in support of this approach will be presented.

We started by adjusting the combination feeder-reflector for maximum gain in the forward direction, and noted that this adjustment remained optimum (within experimental error) even when the row of directors was added. This result has also been established by the Tinkertoy Antenna Group [8]. It was possible to retain the same adjustment of the feeder-reflector combination in successive experiments involving different rows of directors.

If we consider the wave traveling along the array as a surface wave propagating with a certain phase velocity —an approach first adopted by Stanford Research Institute [9]—we can show that the maximum gain of a Yagi antenna of given length L occurs at a definite value of the phase velocity v_x. This value is a function of element spacing s_D, element height h_D, and element radius ρ_D, and a criterion has thus been found for the optimum combination of these parameters.

The separate effects of the two parts of a Yagi antenna are shown in Figs. 3 and 4. These figures are phase and amplitude plots of the nearfield of a Yagi 6 λ long and are adjusted for maximum gain. The patterns were obtained with the automatic plotter built at the Air Force Cambridge Research Center [10]. In Fig. 3 the reflector has been removed; in Fig. 4 the field of the entire Yagi is plotted. We note that the phase lines lying perpendicular to the row of directors are undisturbed by the addition of the reflector, thus proving that one and the same surface wave is generated. The shift in corresponding amplitude lines A, B, C, D (plotted at 5-db intervals) show that in the vicinity of the feeder, the pattern is omnidirectional in the absence of the reflector, and directive when the reflector has been added.

Experimental Arrangement

Because it was necessary to have a setup in which the different parameters of the Yagi could easily be changed, it was decided to conduct the experiments above a ground plane. Another obvious advantage of using a ground plane was that the feeder element could be excited from below, leaving the antenna unobstructed. The individual radiators were brass rods, and their height was varied by press-fitting the rods in a set of holes drilled in the ground plane. Mechanical support of the elements was furnished by the ground plane. The spacing between elements could easily be changed in multiples of the smallest separation.

To measure the pattern and gain in the farfield of the antenna, the ground plane had to be extended far out. We decided to work at X band (3.3 cm) so that the dimensions could be kept within reasonable limits. It was sufficient to let the ground plane extend to the far zone of the antenna in one direction only. We chose a width of 6 feet and a length of 12 feet (110 wavelengths).

The rods were set in holes drilled in a circular plate which was recessed and free to rotate in the ground plane. To plot the radiation pattern of the Yagi, it was only necessary to rotate this circular plate while the receiving antenna was held fixed. The following appear in Fig. 5, which shows the diagram of the experimental setup used in all measurements.

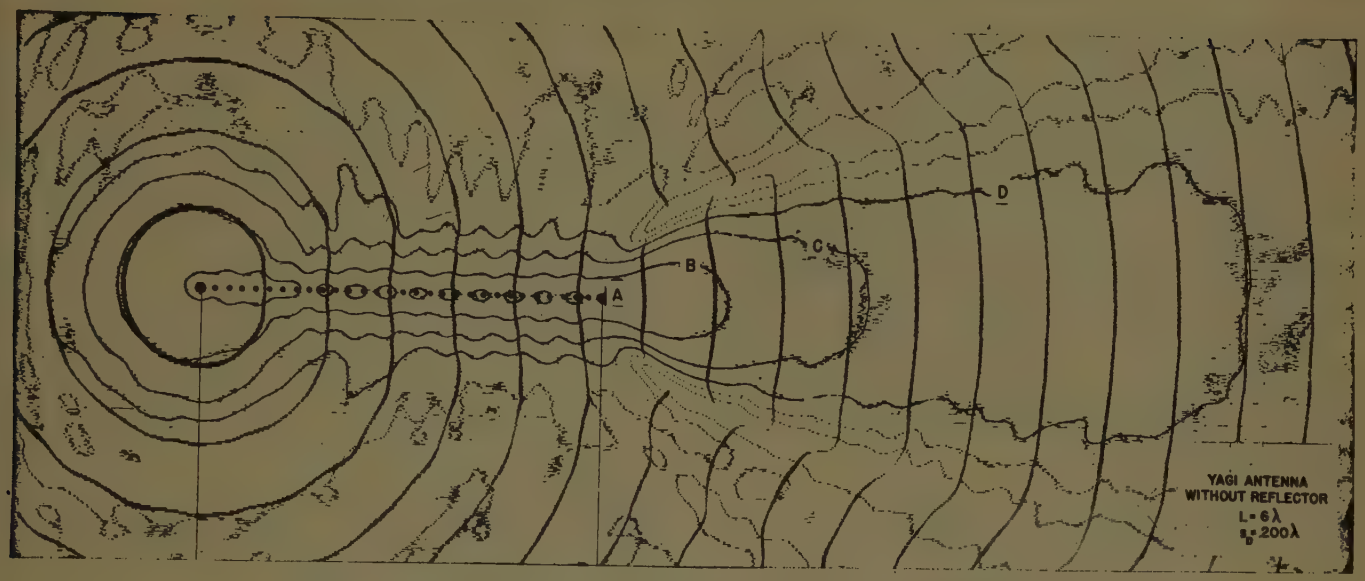


Fig. 3—Nearfield amplitude and phase plot of Yagi antenna without reflector.

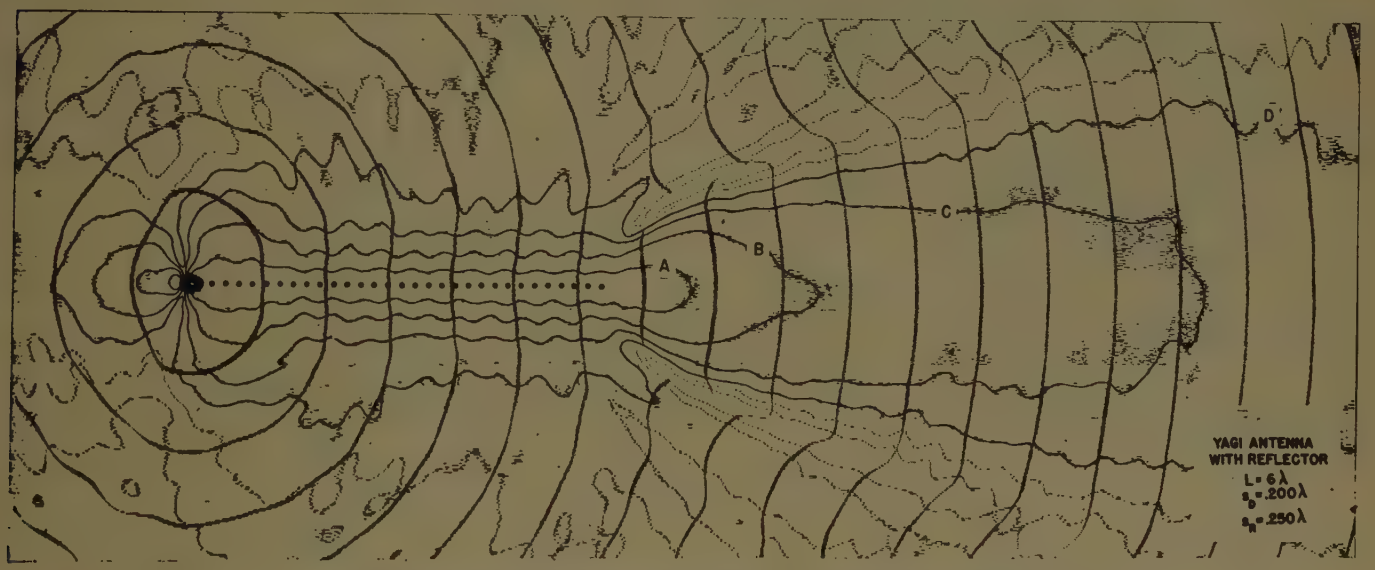


Fig. 4—Nearfield amplitude and phase plot of complete Yagi antenna.

Y1 the audio oscillator (1000 cps) used with MOD.
Y2 the frequency-stabilized source operating at 9084 Mcps.
MOD the ferrite-type modulator for amplitude modulation of the signal source.
AT1 the attenuator that measures the antenna gain. [The power input to the antenna was attenuated until the field strength in the farfield was the same as obtained from the reference monopole (the feeder).]
ST1 the standing-wave indicator.
TR1 the impedance transformer.
MT1 the mode transducer from waveguide to coaxial line.
ANI the receiving antenna in the farfield. (The detector monitors the signal strength.)

Fig. 5 also shows the method used for measuring the phase velocity along the row of directors. A probe was

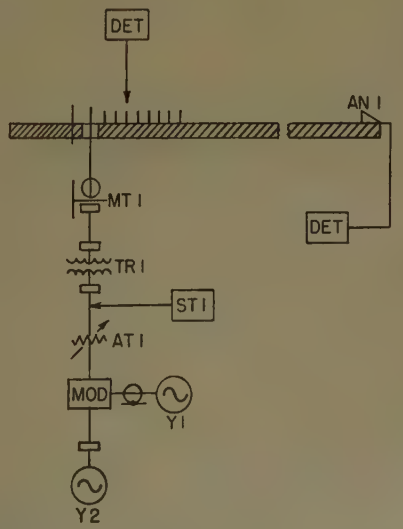


Fig. 5—Block diagram of experimental setup.

carried along a line parallel to the array, at a small distance from the tip of the brass rods. The phase velocity was then determined by measuring the distance between minima. The standing wave along the array was enhanced by placing a metal reflector plate at the end of the Yagi.

The results of the measurements are described in the following section. All experiments were conducted in a reflectionless room, the ground plane also surrounded with absorbing material to reduce reflections to a minimum. All mechanical parts were machined to a tolerance of less than 0.001 inch. Measurements taken on different days could be repeated with a relative error of no more than ± 0.1 db.

Experiments on the Properties of Yagi Antennas

Optimum Adjustment of the Combination Feeder-Reflector

To obtain optimum design criteria for reflector height and spacing, the power in the farfield of the combination feeder-reflector was measured as a function of reflector height and spacing. The height of the feeder, adjusted for resonance, and the input power were kept constant.

Fig. 6 shows the result of these measurements for reflector spacings between 0.083 λ and 0.313 λ, with $\rho_R/\lambda = 0.024$. The maxima of these curves, plotted as a function of reflector spacing (see Fig. 7), show that the adjustment of the combination feeder-reflector for maximum gain is not too critical, although there is an optimum value for the spacing 0.250 λ. With this spacing, and with a reflector height of 0.226 λ, the gain of the combination feeder-reflector is 3.85 db above the feeder alone.

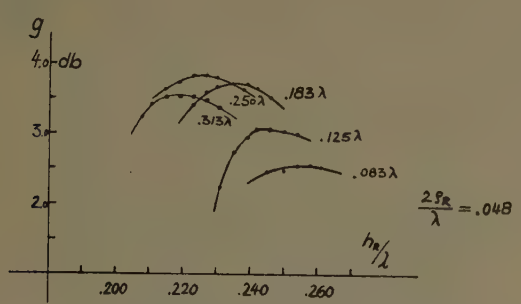


Fig. 6—Gain of feeder-reflector combination as a function of reflector height for various reflector spacings.

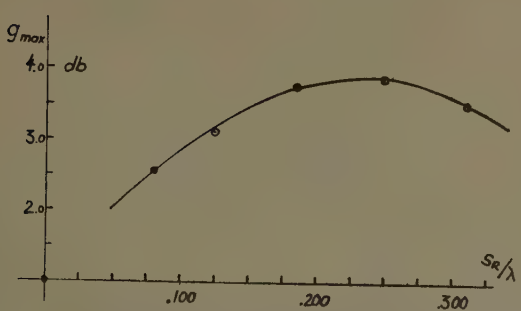


Fig. 7—Maximum gain of feeder-reflector combination as a function of reflector spacing.

Optimum Adjustment of the Directors

Just as in the feeder-reflector case, it is possible to determine the maximum gain as a function of the director parameters but more difficult to obtain a general picture since at least four parameters are involved (s_D, h_D, ρ_D and L). In displaying our experimental results we are restricting ourselves to the more interesting examples, bearing in mind that we are chiefly interested in showing the effect of the parameters on one characteristic, the phase velocity along the row of directors.

In Fig. 8 gain was plotted against director height for several different spacings, for two different spacings, and for two different lengths ($L = 1.2$ λ and $L = 6.0$ λ). The radius ρ_D was held constant throughout. The maximum obtained from these curves was plotted against director spacing in Fig. 9. We note that up to spacings of 0.3 λ, the maximum gain is independent of spacing and, thus, depends only on the array length and on ρ_D. We are omitting experimental data indicating the further conclusion that within certain limits the maximum

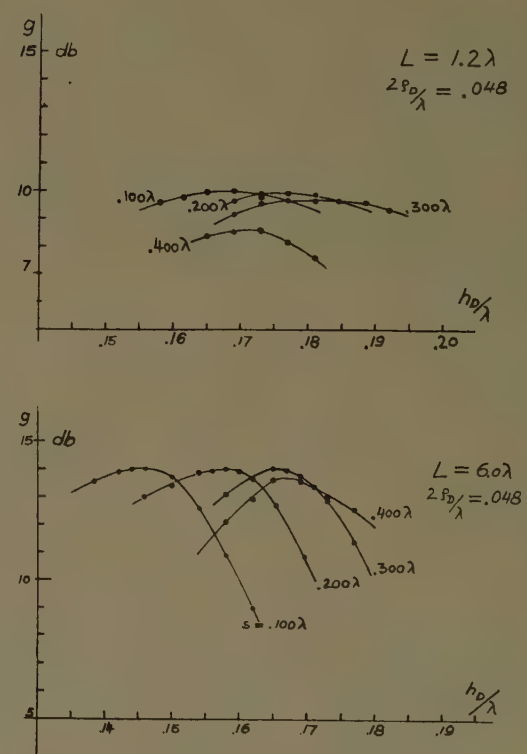


Fig. 8—Gain of Yagi array as a function of director height and spacing for array length.

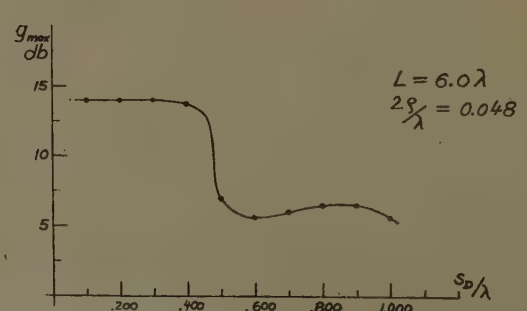


Fig. 9—Maximum gain of Yagi array as a function of director spacing.

gain is also independent of ρ_D, at least up to values of $\rho_D/\lambda = 0.024$.

We noticed that the slight drop in gain for spacings larger than 0.3 λ could be compensated for by inserting an additional director spaced 0.1 λ from the feeder. This director increased the coupling between the feeder and the row of directors. From curves such as those of Fig. 9, but for a sequence of different lengths L, we plotted the solid curve in Fig. 10, showing maximum gain as a function of array length. The broken curve in Fig. 10, showing the relation between maximum gain and L for improved endfire arrays, was taken from Hansen and Woodyard [11]. The difference between the two curves in discussed in the following section.

From the relationship between maximum gain and array length, Hansen and Woodyard deduced the value of the phase velocity traveling along the array and discovered that it was slower than the freespace-propagation velocity by an amount that decreased with increasing length of the array. Since the Yagi antenna differs from an improved endfire array of the Hansen and Woodyard type in its amplitude distribution, we did not know whether we would be able to measure a unique phase velocity. Our experimental results show that such a phase velocity does exist and that it is constant along the row of directors, with small perturbations in the vicinity of each element. Fig. 11 gives the phase velocity, under conditions of maximum gain, as a function of the array length L. Since this phase velocity depends on the parameters s_D, h_D, and ρ_D, it is easier to consider maximum gain as a function of phase velocity than as a function of three individual parameters. Once we know how the phase velocity depends on s_D, h_D, and ρ_D, we can design Yagi antennas by focusing attention on the single parameter instead of on the three constituent parameters. The method by which the optimum phase velocity is obtained is quite unimportant since infinitely many combinations of director spacing, height, and diameter yield the same phase velocity. Fig. 12 shows the measured phase velocity as a function of director height, with the spacing as parameter and the diameter held constant. The measurement of phase velocity for short Yagis being difficult, it is advisable to proceed by first determining the director height for maximum gain, and then taking the phase velocity from the values plotted in Fig. 12.

We have already remarked that different array lengths require different phase velocities for optimum gain. It is possible to show the dependence of phase velocity, array length, and maximum gain upon director height, with director spacing and diameter as parameters. We can limit ourselves to spacings of 0.100 λ, 0.200 λ, 0.300 λ and 0.400 λ, and draw four charts that give all the information required for designing a Yagi antenna. Interpolation takes care of intermediate values of spacing. Fig. 13 contains two such charts, for $s_D = 0.200\ \lambda$ and $s_D = 0.400\ \lambda$.

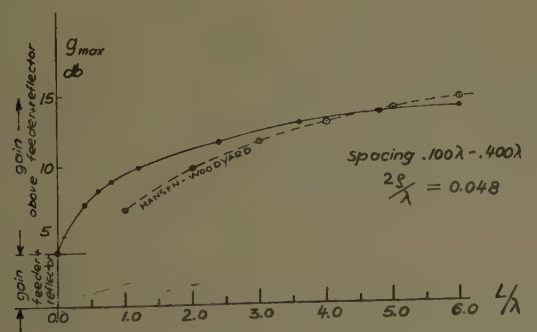


Fig. 10—Maximum gain of Yagi array as a function of array length.

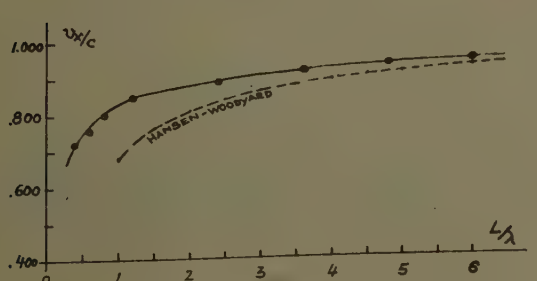


Fig. 11—Phase velocity for obtaining maximum gain as a function of array length.

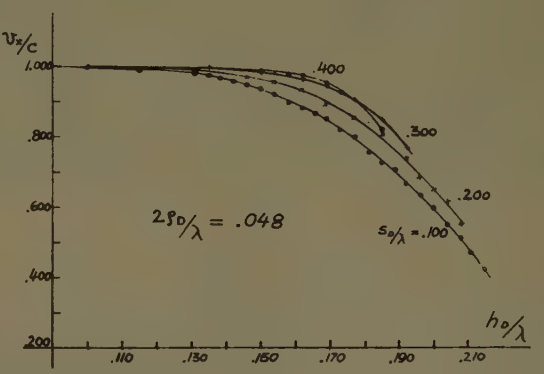


Fig. 12—Phase velocity as a function of director height.

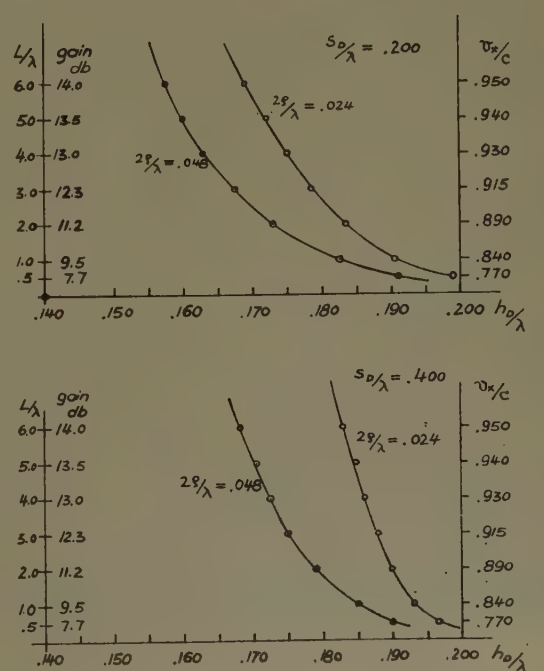


Fig. 13—Gain, phase velocity and array length of an optimized Yagi array as a function of director height.

Comparison with the Hansen-Woodyard Condition

The introduction of phase velocity as the criterion for the design of a maximum-gain Yagi antenna invites a comparison with the Hansen-Woodyard condition. In Fig. 11 we have already shown the experimental discrepancy between our case and theirs. In their case, Hansen and Woodyard [11] assumed that the current distribution along the array is constant, that the element spacing is infinitesimal, and that the array is long compared with the wavelength. In the case of the Yagi antenna, their first assumption is clearly not applicable, as can be seen from our Fig. 14, which gives the current distribution measured along a Yagi array of length 3 λ, adjusted for maximum gain.

Let us consider the field strength in the farfield as a superposition of the field strengths produced by the individual array elements. The vector diagrams of an endfire array whose current distribution is constant along the array and of a Yagi antenna whose current distribution varies as shown in Fig. 14 are, respectively, shown in Figs. 15 and 16 for the endfire direction. In both figures the array length is taken as 3 λ. With an element spacing of 0.3 λ, there are consequently eleven elements in each array.

To fulfill the condition for maximum gain in the Hansen-Woodyard case (see Fig. 15), we need to adjust the director height such that the resulting phase velocity is given by

$$\frac{v_x}{c} = \frac{L/\lambda}{L/\lambda + 0.468}.$$

This implies that the phase delay from element to element has a constant value of 16.8°. The vector diagram of this array is nearly a semicircle, the total phase delay being close to 180°.

The vector diagram in Fig. 16 is for a Yagi with the same phase delay of 16.8° but with the current amplitudes obtained from the measured curve in Fig. 14. We note that the vector sum of the eleven elements is smaller than the maximum obtainable from this array. The vector reaches its maximum length at a point along the curve that corresponds to a total phase delay considerably smaller than the value given by Hansen-Woodyard. This phase delay, which totals 118°, is reached after the contributions of only eight elements have been added. If we wish to obtain the maximum length of the total field vector after eleven elements, we would have to impose a smaller phase delay from one element to the next. Fig. 17, a vector diagram for such an eleven-element Yagi, shows the difference between this case and the ordinary constant amplitude case of Fig. 15. From Figs. 16 and 17 it follows that to obtain maximum gain on this Yagi array the phase velocity must be adjusted to a value of 0.902 times the freespace velocity. The experimental value obtained from Fig. 11

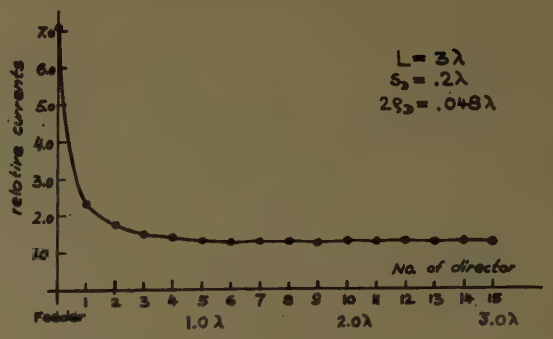


Fig. 14—Current distribution in optimized Yagi array of length 3λ.

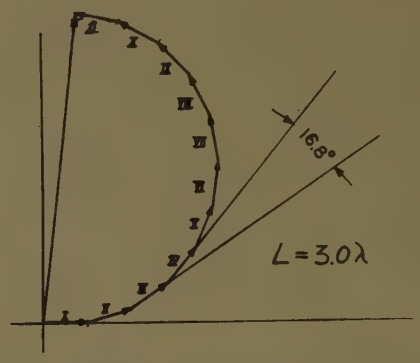


Fig. 15—Vector diagram of field strength in the endfire direction of an array having a constant current distribution.

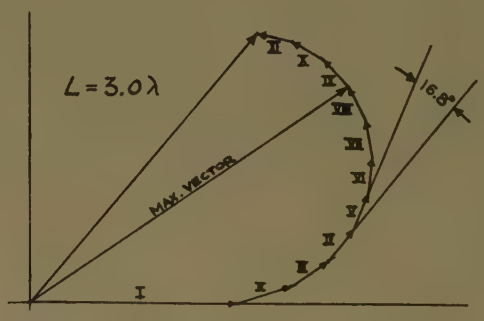


Fig. 16—Vector diagram of field strength in the endfire direction of a Yagi array whose current distribution varies as shown in Fig. 14.

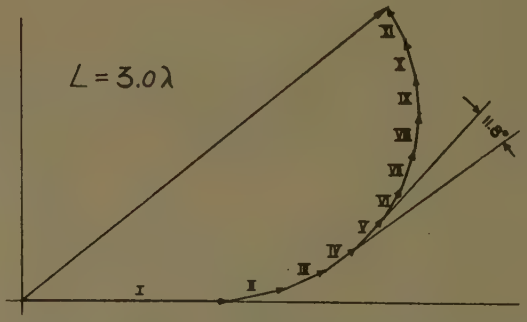


Fig. 17—Vector diagram of field strength in the endfire direction of a Yagi array whose current distribution varies as shown in Fig. 14, with the array adjusted so that the endfire field strength is maximized.

is 0.905 times the freespace velocity, and thus within three per cent of the value given by the graphical method.

The assumption made by Fishenden and Wiblin [5] that the vector diagram of a Yagi should be nearly a semicircle is manifestly incorrect because the current distribution along the array is not, as they supposed, constant. For the same reason, Reid's theoretical results [4] are not applicable either.

To avoid confusion, it should be pointed out that the graphical representation of the farfield intensity enables us to locate the vector position for maximum gain within a given diagram but does not allow comparison between resultant vector magnitudes of different diagrams since the scale vector of each diagram is left undetermined.

Design Method for Yagi Antennas

The optimum design procedure for a Yagi antenna depends on the conditions that must be met. In the most usual situation, the requirement is either for a Yagi with a certain gain in the forward direction or for a Yagi of fixed length and maximum gain.

Let us assume that we wish to design a Yagi over a ground plane such as the fuselage of an airplane. Fig. 13 gives all the information we need. We find in this figure the maximum gain that can be obtained, and also the height of the director elements that must be chosen for a specified spacing. If the element diameter differs from those assumed in Fig. 13, then additional curves must be plotted in, obtained either by interpolation or through experiment.

If the director elements have some shape other than cylindric or if the Yagi is of the conventional type, with a row of dipoles supported on a boom rather than a row of monopoles over a ground plane, then the design procedure is as follows. We again refer to Fig. 13 but ignore the curves and fix attention only on the relation between maximum gain, array length, and phase velocity as read directly from the three scales along the ordinate. Given the gain or given the fixed length of the array, we immediately determine the phase velocity that is required. We must now perform the experimental work of adjusting the director cross section, spacing, and height, so as to obtain the required phase velocity. Mechanical considerations will usually fix one or two of these parameters and so no more than one or two series of data need be taken. (Note that our design method for maximum gain requires no farfield measurements.) It is not necessary to test the phase velocity on the full length array; if the array is at least 3 λ long, the phase velocity will remain unchanged as further elements are added.

As already mentioned, we can choose any spacing less than 0.5 λ and still obtain the maximum gain. If, for example, we are designing a Yagi in the microwave region, we might decide to insert rods at fixed distances along a metal strip. For mechanical stability we would choose very close spacing. If, on the other hand, we are designing a Yagi for lower frequencies—for example, to launch a ground wave—then the number of elements would have to be as small as possible to conserve material costs, and a suitable spacing would be between 0.40 λ and 0.50 λ. Figs. 18 and 19 show these two examples of Yagi antennas. The first director in Fig. 19 is an element that increases the coupling between the feeder and the rest of the directors, as described previously.

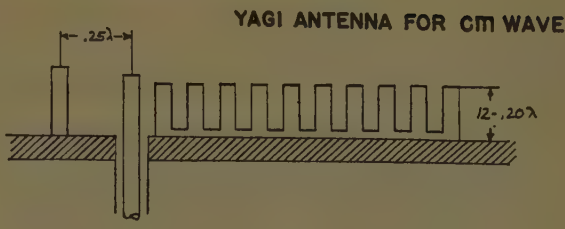


Fig. 18—Yagi array for very high frequencies.

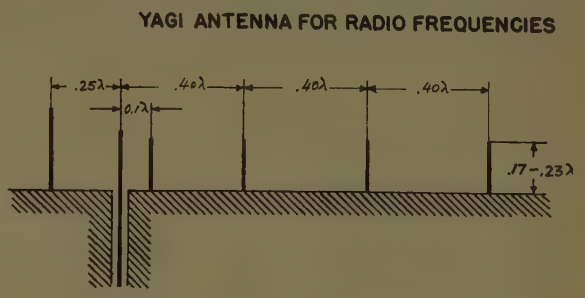


Fig. 19—Yagi array for low frequencies.

Conclusion

We have shown in this report that the phase velocity of the surface wave traveling along the row of directors in a Yagi antenna can be used as the design criterion for maximum gain.

We have also shown that the Hansen-Woodyard condition is inapplicable to the Yagi antenna because the amplitude distribution is not constant along the array. Figs. 10 and 11 give the values of the maximum obtainable gain and the corresponding optimum phase velocity on the Yagi antenna.

To date, very long arrays of the Yagi type have been successfully designed only by J. C. Simon [12], who uses periodic variations in the spacing and height of elements. This procedure has not yet been explained theoretically, and no optimum design procedures are available. It is clear, however, that the explanation will again have to come from a consideration of the spectrum of surface waves propagating along the array.

Acknowledgments

This work was done under the general direction of F. J. Zucker, chief of the Airborne Antenna Section, and R. E. Hiatt, chief of the Antenna Laboratory. The authors are indebted to Dr. W. Gerbes and F. J. Zucker for many stimulating discussions, and to W. J. Kearns for assistance in making the measurements.

References

[1] H. Yagi, "Beam transmission of ultra short waves," Proc. IRE, vol. 16, pp. 715–741; June, 1928.
[2] S. Uda and Y. Mushiake, "Yagi-Uda Antenna," Sasaki Co., Ltd., Japan; 1954.
[3] W. Walkinshaw, "Treatment of short Yagi aerials," *J.IEE, (London)*, vol. 93, pt. 3A, no. 3, pp. 598–614; 1946.
[4] D. G. Reid, "The gain of an idealized Yagi array," *J. IEE, (London)*, vol. 93, pt. 3A, no. 3, pp. 564–566; 1946.
[5] R. M. Fishenden and E. R. Wiblin, "Design of Yagi aerials," *Proc. IEE (London)*, vol. 96, pp. 5–12; March, 1949.
[6] D. M. Vysokovsky, "Amplitude and phase relations of currents in coupled oscillator antennas," *Doklady Akad. Nauk. S.S.S.R.*, vol. 96, pp. 971–974; May, 1954.
[7] D. M. Vysokovsky, "Resonance in a system of coupled oscillator and a tuned antenna," *Doklady Akad. Nauk. S.S.S.R.*, vol. 96, pp. 659–662; April, 1954.
[8] Tinkertoy Antenna Group, Servo Corp. of America, 2nd Quart. Rept., Contract DA 36-039SC-64689; December 15, 1955.
[9] "Ridge and Corrugated Antenna Studies," Stanford Res. Inst., Stanford, Calif., Final Rept., Contract W 36-039SC-44524, Signal Corps Project No. 22-122B-O; December, 1950.
[10] R. M. Barrett and M. H. Barnes, "Automatic antenna wavefront plotter," *Electronics*, vol. 25, pp. 120–127; January, 1952.
[11] W. W. Hansen and J. R. Woodyard, "A new principle in directional antenna design," Proc. IRE, vol. 26, pp. 333–345; March, 1938.
[12] J. C. Simon and V. Biggi, "Un nouveau type d'aerien et son application à la transmission de television a grande distance," *L'Onde Electrique*, vol. 34, pp. 883–896; November, 1954.

A Dipole Antenna Coupled Electromagnetically to a Two-Wire Transmission Line*

S. R. SESHADRI† and K. IIZUKA†

Summary—**The properties of a dipole antenna coupled electromagnetically to a two-wire transmission line are studied experimentally. It is found that the coupling of the antenna to the transmission line can be maximized by a proper choice of 1) the angular position of the antenna with respect to the transmission line, 2) the length of the antenna, and 3) the separation of the antenna from the transmission line. The effect of the spacing between the wires of the transmission line on the optimum parameters is investigated. It is found that the optimum angular position of the antenna is not noticeably altered if, instead of a single antenna, an array of properly located antennas is used as the load. The advantage of an antenna array built on this coupling principle is discussed.**

Introduction

FOR some high-power VHF-UHF communications purposes, it is desirable to have a high-gain linear array whose elements are driven by voltages of proper amplitude and phase. This arrangement involves several transmission lines which cannot all be in the neutral planes of all the others. The unavoidable coupling between the lines and between the antennas and the lines makes the realization of a predetermined pattern difficult, if not impossible. Therefore, it is desirable to reduce the number of transmission lines.

In the region of very short wavelengths, arrays consisting of several slots on a waveguide wall are in common use. An analog of the slot array driven by a single waveguide is an array of dipoles driven by a single transmission line. Such an arrangement was first proposed by Sletten.[1] It consists of several dipoles situated in a plane parallel to that of a two-wire transmission line with their centers adjacent to the positions of voltage maxima on the line. When the axis of the dipole is parallel to the direction of the line the antenna is not excited, since it is in the neutral plane. However, when the antenna is rotated, it is excited in an amount that depends, among other factors, on its angular position. Consequently, the several antennas of the array may be excited with currents that differ from one another. Note that the elements of this array can all be driven from a single two-wire line with currents of predetermined amplitudes. As compared to the waveguide slot array, greater power can be obtained from this dipole array. Also, this array of dipoles coupled electromagnetically to a two-wire line has a desirable radiation pattern that is useful for communications purposes.

In order to understand and design a practical array of dipoles, it is first necessary to know the properties of a single antenna coupled to a two-wire line. It is therefore the purpose of this research to investigate the circuit properties of a dipole antenna coupled electromagnetically to a two-wire transmission line. The field properties are reserved for a later study.

* Manuscript received by the PGAP January 19, 1959; revised manuscript received, April 27, 1959. This research was performed under Air Force Contract No. AF19(604)-786.

† Gordon McKay Laboratory of Appl. Phys., Harvard University, Cambridge, Mass.

[1] C. J. Sletten and G. R. Forbes, "A New Type of Antenna for VHF-UHF Communications," Air Froce Cambridge Res. Center, Bedford, Mass., Tech. Rept. ERD-CRRD-TM-56-103; January, 1956.

Description of the Experimental Arrangement and the Measuring Technique

The experimental arrangement used in this investigation and the technique of measurement are well known so that only a brief description is required. Details may be found in Seshadri and Iizuka.[2]

The apparatus consists of a two-wire transmission line seven wavelengths long, made of brass wires 0.32 cm in diameter. The spacing b between the wires can be adjusted to one of the following values: $b = 1.26$, 2.21, 3.16 and 4.11 cm. The transmission line is short-circuited at the load end. There is a tandem bridge at the generator end that may be moved to tune the line to sharp resonance. The line is excited in such a manner that in the absence of the load only "balanced currents"[3] are present. However, in the presence of the load, "unbalanced currents"[3] are also excited since the load is unsymmetrical with respect to the two wires.

A schematic diagram of the transmission line with its exciting, monitoring, and detecting systems is shown in Fig. 1. The input circuit is arranged to excite the line in

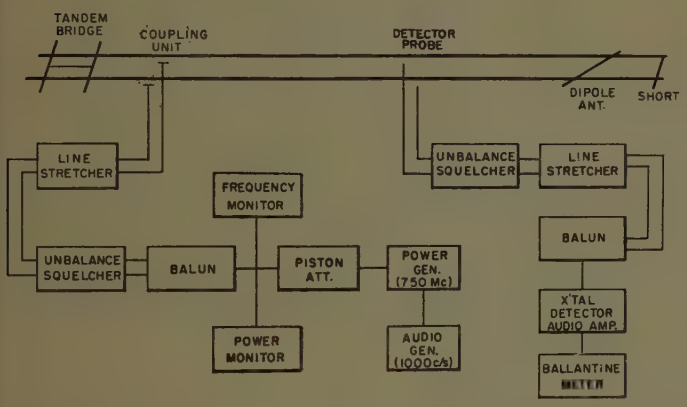


Fig. 1.

the balanced mode in the absence of the load. The detector circuit measures the balanced component of the current on the line.

An electromagnetically-coupled dipole antenna serves as the load. It is possible to adjust its angular position θ with respect to the transmission line (see Fig. 2) and to align its center to lie midway between the two wires. The distance d between the antenna and the plane of the transmission line can be adjusted to any desired value. The length $2h$ of the antenna may be varied from $2h = 16$ cm to $2h = 24$ cm. At the frequency at which this experiment was performed, $2h = 20$ corresponds exactly to a half wavelength.

[2] S. R. Seshadri, and K. Iizuka, "A Dipole Antenna Coupled Electromagnetically to a Two-Wire Transmission Line," Cruft Laboratory, Harvard University, Scientific Rept. No. 21; 1958.

[3] R. W. P. King, "Transmission Line Theory," McGraw-Hill Book Co., Inc., New York, N. Y.; 1955.

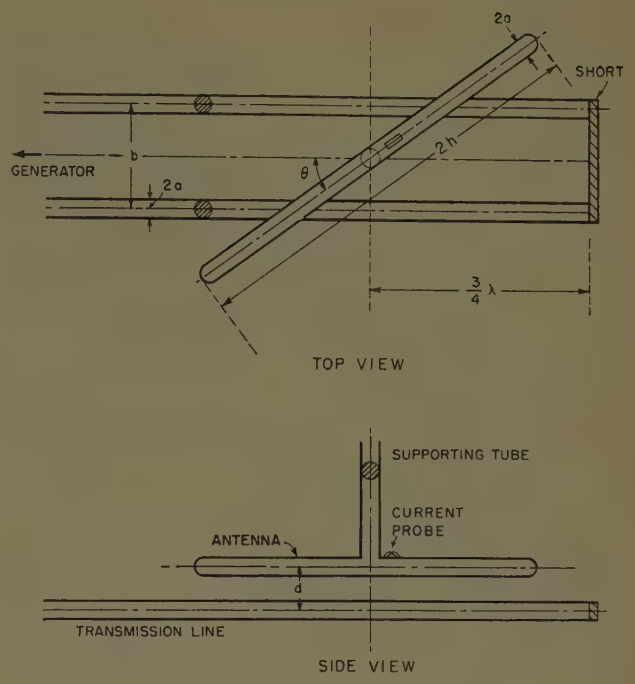


Fig. 2.

The principal measurements were the determination of the standing-wave ratio on the transmission line for various antenna configurations. In order to get results which could be compared, it was found necessary to take a whole series of measurements in one run. In cases where the equivalent impedance was desired, the position of the voltage minimum had to be measured. Since its position was sharp in most cases, its position could be located with accuracy. The entire system was first calibrated and all meter readings corrected.

Coupling as a Function of Angle of Rotation

The distribution of voltage along the transmission line is first measured and the antenna is placed so that its center corresponds to the position of voltage maximum. The antenna is then adjusted to lie midway between the two wires, and, to begin with, is aligned parallel to them. Since it is in the neutral plane, the pattern of the standing waves on the line is not disturbed. If the antenna is rotated about a vertical axis, however, it is excited and radiates. In order to investigate the dependence of the coupling of the antenna on the angle of rotation, a suitable parameter capable of describing this coupling is required.

When the antenna is in a position in which it couples appreciable power from the transmission line, it is asymmetrical with respect to the two parallel wires. Unbalanced currents are, therefore, reflected back by the load on the line. Naturally, the question of a parameter suitable for describing the coupling of the antenna to the transmission line arises. Roughly speaking, the load can

be treated in two parts of which the first reflects only balanced currents and may be referred to as the symmetrical part of the load. It may be represented by an equivalent impedance shunting the transmission line at a convenient reference point, for example, the position on the transmission line just beneath the center of the antenna. This equivalent impedance can easily be ascertained by making the usual standing-wave measurements on the line. The relative position of a voltage minimum and the magnitude of the standing-wave ratio completely determine this impedance. However, the magnitude of the inverse of the standing-wave ratio is a more suitable parameter characterizing the symmetrical part of the load. For this part of the load, the only relevant currents on the transmission line are balanced currents. Since the radiation from such a transmission line is small, the difference between the incident and the reflected power may be attributed to the radiation from the antenna. It can easily be shown that $(P_i - P_r)/P_i = 4s/(1+s)^2$, where P_i is the incident power, P_r is the reflected power, and s is the inverse of the standing-wave ratio on the line. Therefore, the difference between the incident and the reflected power increases with the inverse of the standing-wave ratio. Hence the inverse of the standing-wave ratio can be taken as an adequate measure of the power of that part of the antenna radiation which gives rise to the symmetrical part of the load.

The second part of the load, which is responsible for the unbalance component of the currents on the two wires of the line, may be referred to as the asymmetrical part. A measure of this part of the load is the unbalanced component of the current on the transmission line. When codirectional currents exist on the two-wire line, the radiation from it is comparable in magnitude to that from the antenna. Hence, for this part the entire transmission line together with the antenna constitutes the radiating system. This radiation disturbs the field pattern that would otherwise be obtained. Fortunately, it is found that the relative magnitude of the unbalanced component of the currents on the lines is small. It is therefore reasonable to conclude that the symmetrical part of the load predominates over the asymmetrical part and to assume that the radiation from the two-wire line is small compared to that from the antenna. In a practical antenna system an array of dipoles is located at points corresponding to the positions of successive voltage maxima. For example, in an array of two antennas, one antenna is rotated clockwise away from the direction of the transmission line while the other is rotated counterclockwise through the same angle. It is seen that in such an array the unbalanced component of the currents on the line is still further reduced. In effect, the asymmetrical parts of the loads due to the antennas tend to cancel whereas the symmetrical parts add. Thus it is evident that the symmetrical part of the load plays a more important role. Consequently in the investigation of the variation of coupling with the angle of rotation, the length of the antenna and the separation distance between the antenna and the transmission line, the inverse of the standing-wave ratio may be used as a parameter to describe the effectiveness of coupling.

In order to determine the dependence of the coupling on the angular position of the antenna, an experiment was performed on a dipole antenna of length $2h = 20$ cm at a distance $d = 0.4$ cm. Fig. 3 shows the measured variation of the standing-wave ratio as a function of the angle of rotation θ. This graph shows a marked peak at a particular angle θ_m. This angle is referred to as the optimum angle since for a given antenna length $2h$, and separation d, it gives the angle at which maximum power is radiated from the antenna. Within the limits of experimental error, the graphs are symmetrical about $\theta = 0$ since in the absence of the load the transmission line is excited in the balanced mode. The symmetry of the graphs about $\theta = 0$ actually serves as a check on 1) the balanced excitation of the line in the absence of the load, and 2) the exact placement of the antenna above the center of the two-wire lines.

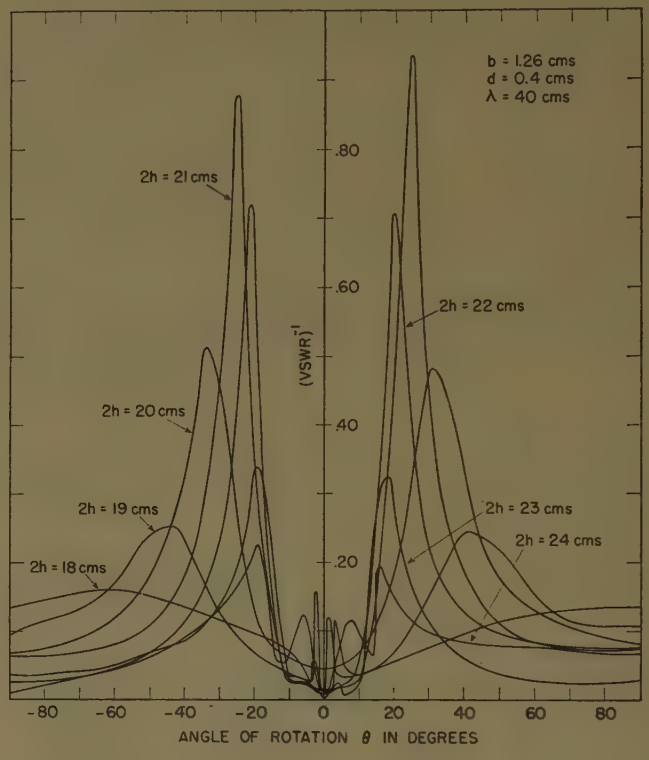


Fig. 3.

A second peak is also shown in Fig. 3. It appears well defined for some values of $2h$ and occurs at an angle smaller than the optimum that corresponds approximately to the position in which the ends of the antenna are just above the two wires. It is therefore conjectured that these subsidiary peaks may be due to voltages induced near the ends of the antenna by the charges on the transmission line.

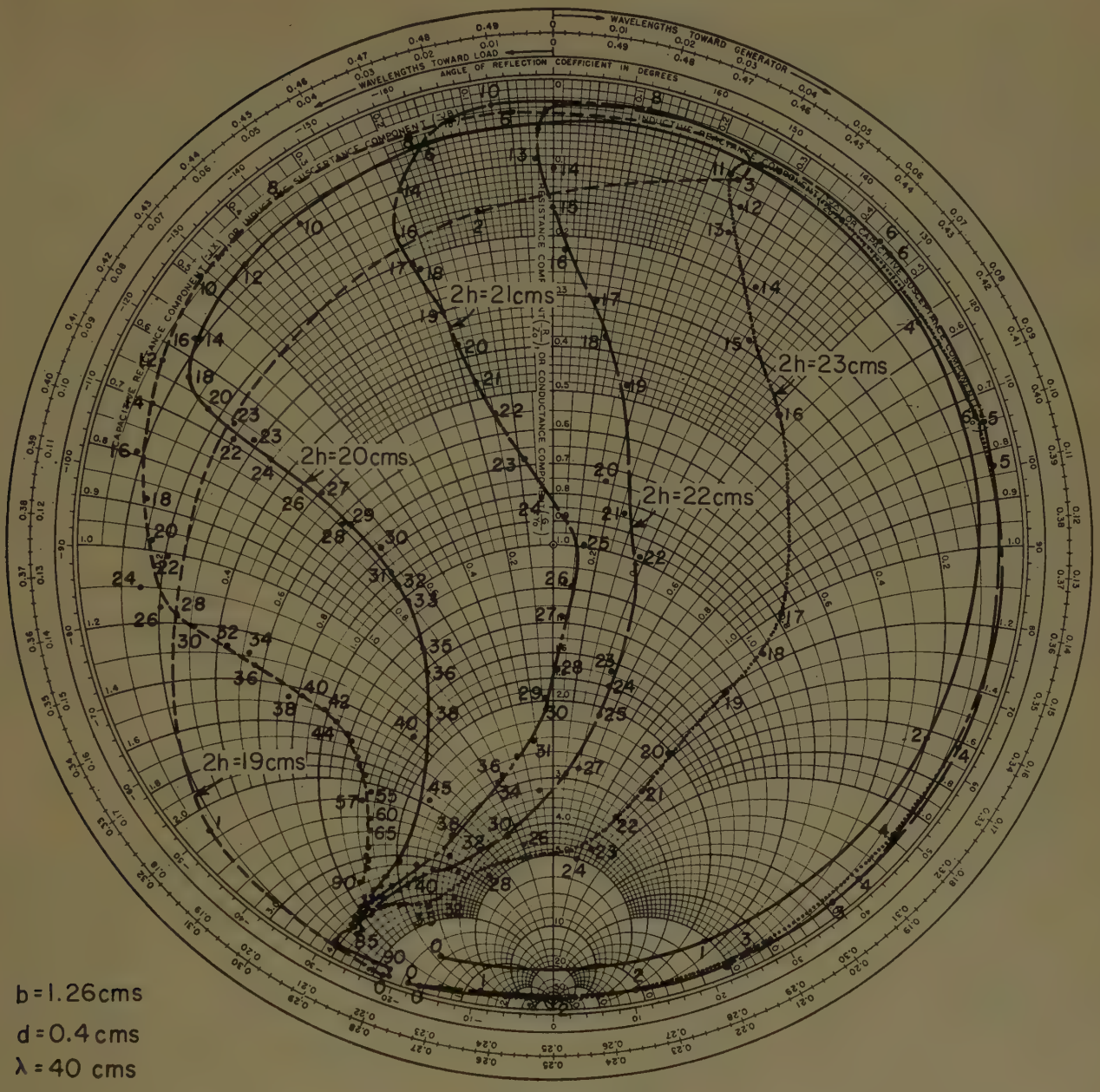


Fig. 4.

The normalized equivalent impedance that loads the transmission line at a reference position corresponding to the center of the antenna is shown in Fig. 4. It is to be noted that the technique used is not very accurate for high standing-wave ratios. Hence the points on the Smith chart (see Fig. 4) which correspond to large standing-wave ratios are not as accurate as those which correspond to smaller values of the standing-wave ratio. The purpose of the experiment was to find the angular position corresponding to maximum loading which occurs at small values of the standing-wave ratio. Hence the lack of accuracy in the measurement of high standing-wave ratios does not limit the accuracy or the scope of the experiment. When the antenna is parallel to the transmission line, it is not excited and the corresponding load impedance should be that for an open circuit; but the experiment shows a departure from this expected value. This is a consequence of the lack of accuracy of the experiment in high values of standing-wave ratio. All of the impedance plots show this kind of departure from the expected form, and they must be similarly interpreted.

It is seen from the impedance graph (see Fig. 4) that the reactance varies rapidly in about the first ten degrees of rotation, whereas the resistance in this range remains almost unaltered and large, indicating very little radiation of power from the antenna. The coupling of power from the transmission line to the antenna oc-

curs at greater angles; it increases with the angle of rotation. At an angle which corresponds to the point on the curve closest to the center of the chart, the coupling becomes a maximum. If the antenna is rotated further, the coupling becomes smaller until it finally becomes negligible.

It is pertinent to note that the reactive part of the equivalent load impedance has both inductive and capacitive values in the entire range of its angular position. This property seems to depend on the length of the antenna, as will be seen later when the impedance graphs for antennas of different lengths are considered.

Effect of Length of Antenna

The loading effect of the dipole antenna depends on its length. In order to study the nature of this dependence, the standing-wave ratio and the position of the voltage minimum were measured for antenna lengths varying between $2h=24$ cm and $2h=16$ cm as a function of the angle of rotation of the antenna. For antennas of lengths 16 cm and 17 cm, the loading effect was small and no pronounced maximum was noticed for any angle of rotation. The plots of the inverse of the standing-wave ratio on the two-wire line as a function of the angle of rotation are shown for the antennas of lengths varying from $2h=24$ cm to $2h=18$ cm in Fig. 3, along with that for the antenna of length $2h=20$ cm. A study of these graphs reveals the following interesting features. As the length of the antenna is decreased from 24 cm to 21 cm, the maximum value of the inverse of the standing-wave ratio is seen to increase steadily. If the length of the antenna is still further decreased from 21 cm to 18 cm, the maximum value of the inverse of the standing-wave ratio, instead of continuing to increase, shows a steady decrease. It was pointed out that the inverse of the standing-wave ratio on the line is a measure of the coupling of the antenna to the transmission line. Hence, it is reasonable to state that there exists a particular length of the antenna for which optimum coupling occurs. Furthermore, it is to be noticed that the subsidiary peak which is clearly visible for antennas of lengths 23 cm, 24 cm, 19 cm, and 18 cm is not detectable for antennas of lengths closer to a half wavelength.

The equivalent loading impedance of antennas of lengths $2h=23$ cm, 22 cm, 21 cm, and 19 cm is plotted in Fig. 4 together with that for the antenna of length $2h=20$ cm. For antennas longer than 20 cm (half a wavelength) the impedance is largely inductive, whereas it is capacitive for antennas shorter than half a wavelength. It is perhaps interesting to note in this connection that the input impedance of a dipole antenna shows a similar variation with respect to its length; namely, for lengths greater than approximately a half wavelength it is inductive and for shorter lengths it is capacitive.

An examination of the Smith charts on which the equivalent impedances are plotted shows that there is a systematic variation with the length of the antenna of the angular position at which the impedance curves approach closest to the center of the charts. The angular position of the antenna for which the curve is closest to the center of the chart corresponds to the maximum of the inverse of the standing-wave ratio on the line. This optimum angle θ_m is seen to decrease continuously as the length of the antenna is increased. Fig. 5 shows this dependence of the optimum angle on the length of the antenna, not only for the distance $d=0.4$ cm between the antenna and the line, but also for four other distances; $d=0.5$, 0.6, 0.75, and 1.0 cm. For all these distances the same behavior between the optimum angle and the length of the antenna is observed.

A closer study of the Smith charts (see Fig. 4) reveals that the magnitude of the inverse of the standing-wave ratio $(\mathrm{VSWR})_m^{-1}$ corresponding to the optimum angle has a maximum value for a particular length of the antenna. For a distance $d=0.4$ cm the maximum occurs for an antenna lenggh $2h=21$ cm. A similar variation of $(\mathrm{VSWR})_m^{-1}$ with the length of the antenna occurs for other distances d between the antenna and the transmission line, with one noticeable difference. For larger values of d (see Fig. 6) the variation of $(\mathrm{VSWR})_m^{-1}$ with the length of the antenna shows sharper peaks; also as d increases, the peaks occur for antennas of shorter and shorter lengths. However, the magnitude of these peaks is approximately the same.

Dependence on Separation Distance Between Antenna and Transmission Line

It is possible to visualize the antenna as a simple matching device. If the antenna is adjusted so that the equivalent load is very nearly equal to the characteristic impedance of the transmission line, then maximum power is coupled to the antenna. It is then possible to anticipate that the coupling to the antenna need not necessarily be a maximum when it is closest to the two-wire line. This anticipation is borne out by experiment (see Fig. 7). For lengths of antennas varying from 24 to 19 cm, the variation of $(\mathrm{VSWR})_m^{-1}$ with the distance d is shown graphically in Fig. 7 for the transmission line spacing $b=1.26$ cm. The coupling is seen to diminish continuously with an increase in the distance d for antennas of length 21 cm and greater. But for 20- and 20.5-cm antennas, the coupling becomes a maximum for a separation nearly two to three times the diameter of the wire. The optimum separation depends critically on the length of the antenna. For the 21-cm antenna a separation that is equal to the diameter of the wire appears most favorable. But if the length of the antenna is decreased by just 1 cm, that is, 1/40 of a wavelength, the optimum separation becomes approximately equal to three times the wire diameter, or nearly the distance between the wires of the transmission line.

It is also to be noted (see Fig. 8) that the optimum angle θ_m varies with the separation d in a systematic manner. For all antenna lengths, the optimum angle θ_m decreases continuously as the separation is increased; the decrease of θ_m with d is less rapid for long than for short antennas (see Fig. 8).

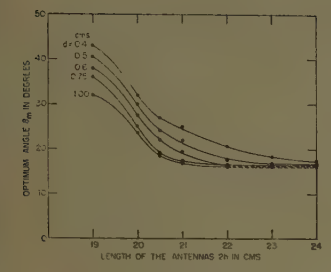


Fig. 5.

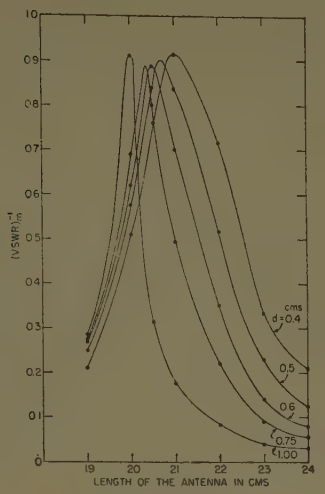


Fig. 6.

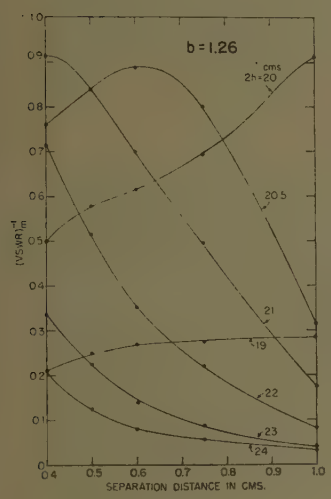


Fig. 7.

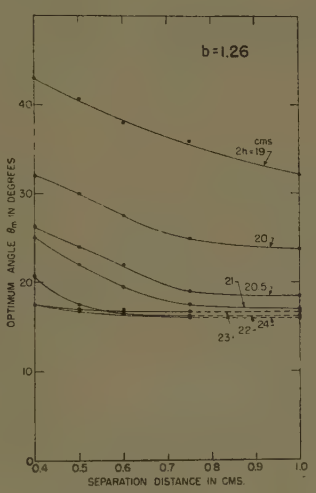


Fig. 8.

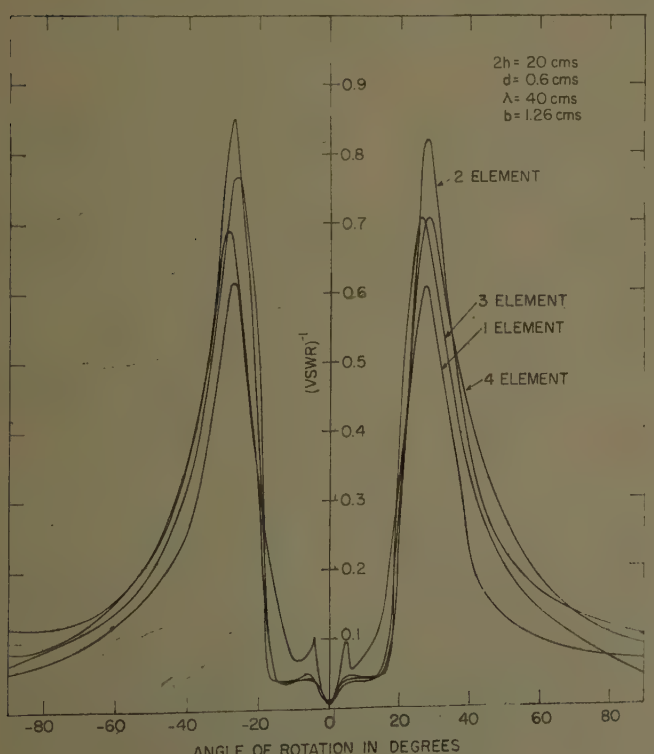


Fig. 9.

Measurement on Arrays of Dipole Antennas

In practical applications a single antenna is not useful; only arrays consisting of 20 to 40 elements give the desired sharpness to the radiated beam. In an array, the antennas are placed at successive voltage maxima and the adjacent antennas are placed in opposite angular positions with respect to each other. This configuration serves two purposes: first, the unbalanced currents on the transmission line are minimized, and, second, a properly polarized beam is obtained.

The knowledge of the optimum parameters related to a single antenna is useful in designing an array only if the values of these parameters, in particular, the optimum angular position of the antenna, are not greatly altered for an array. An experiment was performed on arrays consisting of one to four elements for a fixed separation d and a line spacing b. The results are shown in Fig. 9. It is seen that the optimum value of the angle remains practically the same for all four cases. It may, therefore, be concluded that the coupling between the antennas themselves is small enough not to affect the optimum parameters for coupling of each antenna to the line. Consequently, the values of the optimum angle for single antennas may be used for each element of an array for obtaining maximum coupling effect.

With reference to the design of an array, it is appropriate to make the following comments. Consider a system consisting of several antennas properly located at successive voltage maxima on a two-wire transmission line which is excited at one end. If the element closest to the generator is adjusted to extract maximum power from the line, the other elements of the array draw very small amounts of power. It is therefore necessary to vary the coupling parameters along the array and to keep the coupling of each antenna considerably below its maximum value in order that all the elements in the array may be excited equally. However, it is desirable to keep the angular positions the same for all the antennas except that the adjacent antennas are rotated in opposite directions with respect to each other. This appears necessary for obtaining the proper radiation pattern. The separation d is the parameter to be changed to vary the coupling of the antennas to the line individually. In order to minimize the side-lobe level it is perhaps desirable to taper the excitations towards the end elements; this is not difficult to arrange. It may even be advantageous to excite the line from the center or from both sides, for then the arrangement of the antennas is symmetrical with respect to the center of the array. For the purpose of minimizing the unbalanced currents on the transmission line, it is better to have only even numbers of elements in an array.

Dependence of Optimum Parameters on the Transmission Line Spacing

The spacing b between the transmission lines was kept small (approximately 1/30 of the wavelength) in order that the radiation from the lines would be negligible compared to that from the antenna. In certain cases it

is necessary to increase the line spacing even at the risk of increasing the radiation from the line. When large amounts of power are to be transmitted from the antenna, a large potential difference must be maintained between the two wires of the transmission line. In order to avoid dielectric breakdown, it may become necessary to increase the line spacing. Therefore it is desirable to determine the dependence of the optimum parameters on the transmission-line spacing.

For the spacing $b=2.21$ cm the respective values of $(\mathrm{VSWR})_m^{-1}$ and θ_m are plotted (see Figs. 10 and 11) against the separation d, for different antenna lengths. These curves exhibit the same sort of behavior as those for the smaller line spacing $b=1.26$ cm. The optimum angle θ_m decreases with an increase in the separation d, and the extent of the decrease is greater for short than for long antennas. The variation of $(\mathrm{VSWR})_m^{-1}$ with d (see Fig. 11) for the spacing $b=2.21$ cm is very similar to that for the smaller spacing $b=1.26$ cm, except that this similarity is exhibited for a shorter antenna.

The variations of $(\mathrm{VSWR})_m^{-1}$ and θ_m with the length of the antenna are plotted in Figs. 12 and 13, respectively, for various line spacings. It is seen that the maximum value of $(\mathrm{VSWR})_m^{-1}$ occurs for smaller and smaller antenna lengths as the line spacing is increased. It is also to be noted (see Fig. 13) that the optimum angle θ_m shows a steady decrease as the length of the antenna is increased for a particular separation ($d=0.75$ cm) and for all line spacings.

It is interesting to notice that even as the increase of the line spacing has to be accompanied by a decrease of the length of the antenna in order to have the value of $(\mathrm{VSWR})_m^{-1}$ maximum for a given separation between antenna and line, the increase of the line spacing also has to be accompanied by a decrease in the separation d (see Fig. 14) in order that $(\mathrm{VSWR})_m^{-1}$ may be a maximum for a given length of the antenna. Also, for a given antenna length, the optimum angle θ_m decreases as the separation d is increased for all line spacings (see Fig. 15).

Consequent to the above experimental results on the effect of line spacing, the following observation is useful. If the line spacing is increased to avoid dielectric breakdown, it is necessary to choose the length of the antenna properly in order to avoid the small separations of the antenna and the transmission line that are necessary to obtain the desired coupling to the antenna.

The current at the center of the antenna was measured with a shielded loop probe as a function of the angular position of the antenna with respect to the transmission line. It was found that at the angular position corresponding to which there was a maximum dissipative loading on the transmission line, the current in the antenna was a maximum. Also, the unbalanced currents on the transmission line were measured and were found to be small in comparison with the balanced components. In addition it was observed that the relative magnitude of the unbalanced currents was smaller for an array with an even number than with an odd number of antennas.

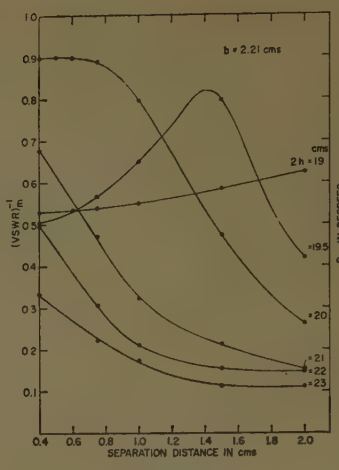


Fig. 10.

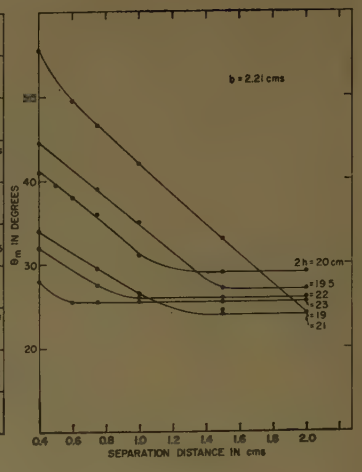


Fig. 11.

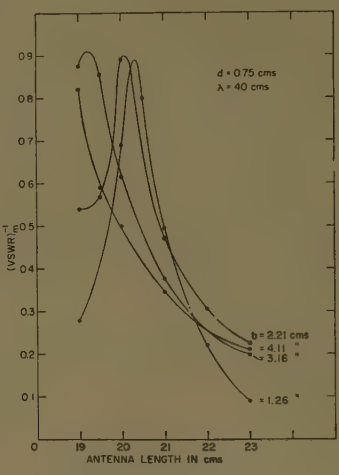


Fig. 12.

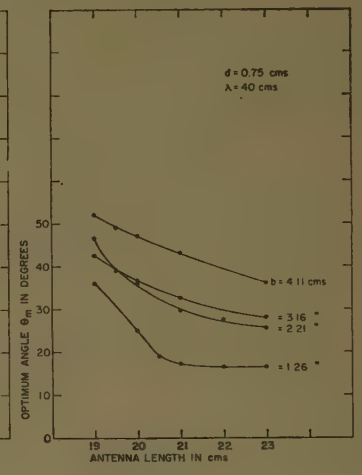


Fig. 13.

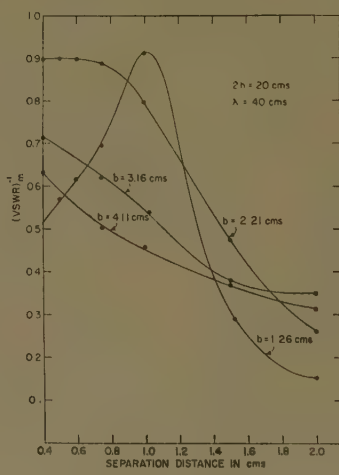


Fig. 14.

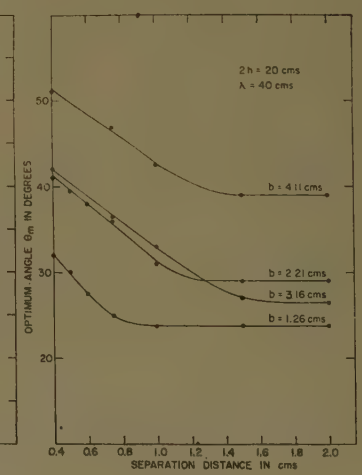


Fig. 15.

Acknowledgment

The authors are indebted to Prof. R. W. P. King for suggesting the problem, for many illuminating discussions in the course of its investigation, and for reading and criticizing the manuscript.

They also wish to thank C. J. Sletten for discussing the problem and Miss Joan Flaherty and Mrs. James Koerner for carrying out the numerical calculations.

An Ionospheric Ray-Tracing Technique and Its Application to a Problem in Long-Distance Radio Propagation*

D. B. MULDREW†

Summary—A method is given for the determination of the equation of a ray path in a known ionosphere where there are no horizontal gradients. It can partially take into account the effects of the magnetic field of the earth. The method was applied to an oblique path between Ottawa and Slough (5300 km) to determine certain properties of the one-hop mode. From this it is shown that at times one hop direct ray propagation is possible over this path.

Introduction

A METHOD for calculating the path of a radio ray in a given ionosphere was proposed by A. H. de Voogt in 1953.[1] In this, a curve of electron density, N, versus distance from the center of the earth, r, is produced from a number of line segments of the form

$$kN = a + \frac{b}{r} + \frac{g}{r^2} \tag{1}$$

where a, b, and g are constants and k is such that $\sqrt{kN}$ is the plasma frequency. For each segment, two of the constants ensure the continuity of the curve and its first derivative at the junction of the previous segment. (For the first segment, N and dN/dr are set equal to zero at the bottom of the ionosphere.) The remaining constant produces the desired distribution. The equations and numerical calculations become successively more complicated with each additional segment. Once the lengthy procedure of finding the equation of the segments has been completed, the expression for the path in each segment can be determined from its differential equation. De Voogt does not go through the complete procedure (because of its complexity), but indicates the steps required to compute the ray path.

In the following method, the electron density distribution is approximated by straight line segments, which are fitted to the distribution by eye; it is a simple matter to determine their equations. It will be shown below that it is not necessary for the first derivatives of the electron density distribution curve to be continuous to ensure that the ray path and its first derivative are continuous. The complete equation for the ray path and for the total angle subtended by the ray path at the center of the earth will be given.

Method for Determining the Ray-Path Equation

For a curved ionosphere with no horizontal gradient, it can be shown, using Snell's law, that

$$n\rho \sin \alpha = \sin \alpha_0, \tag{2}$$

where n is the index of refraction at a distance r from the center of the earth and at a point where the ray path makes an angle α with a straight line through the center of the earth (see Figs. 1 and 3), ρ is the ratio of the distance r to the radius of the earth, and α_0 is the value of α at the earth's surface.

An electron density distribution can be constructed from a vertical ionogram using Schmerling and Thomas's method[2] which neglects the magnetic field, or using Schmerling's method[3] which includes the effects of the magnetic field. The distribution may be represented accurately by fitting to it a number of straight line segments. If N is the electron density at ρ, a_i and b_i are constants, and k is a constant such that $\sqrt{kN}$ is the plasma frequency, then these straight line segments may be approximated by

$$N = \frac{a_i}{k}\rho^2 - \frac{b_i}{k}, \qquad i = 0, 1, 2 \cdots n, \tag{3}$$

since the curvature of (3) is very small for values of ρ applicable to the ionosphere.

The electron density, N, and hence the index of refraction, are continuous functions of ρ throughout the distribution. As can be seen from (2), this makes α a continuous function of ρ which ensures the continuity of the first derivative of the ray path, $d\rho/d\theta$, at all points. Thus, as mentioned above, the ray path and its first derivatives are continuous.

Neglecting collision effects, the index of refraction in the ionosphere,[4] is given by

$$n^2 = 1 - \frac{kN}{f^2 - \dfrac{C_L^2 f^2}{2(f^2 - kN)} \pm \sqrt{\dfrac{C_T^4 f^4}{4(f^2 - kN)^2} + C_L^2 f^2}} \tag{4}$$

where

$$k = \frac{e^2}{\pi m}, \tag{5}$$

$$C_T = \frac{eH_T}{2\pi mC}, \tag{6}$$

$$C_L = \frac{H_L}{2\pi mC}. \tag{7}$$

* Manuscript received by the PGAP, Feburary 2, 1959; revised manuscript received July 9, 1959.

† Defence Res. Telecommun. Estab., Ottawa, Canada.

[1] A. H. de Voogt, "The calculation of the path of a radio-ray in a given ionsphere," PROC. IRE, vol. 41, pp. 1183–1186; September, 1953.

[2] E. R. Schmerling and J. O. Thomas, "The distribution of electrons in the undisturbed F2 layer of the ionosphere." *Phil. Trans. Roy. Soc. London*, vol. 248, pp. 609–620; 1956.

[3] E. R. Schmerling. "An easily applied method for the reduction of h'f records to N-h profiles including the effects of the earth's magnetic field." *J. Atmospheric and Terrest. Phys.*, vol. 12, pp. 8–16; 1958.

[4] S. K. Mitra, "The Upper Atmosphere," The Asiatic Society, Calcutta, India, pp. 187–216; 1952.

The frequency (cycles/sec) is f, e is the electronic charge (esu), m is the electronic mass (gm), c is the velocity of light in free space (cm/sec), H_T is the component of the magnetic field transverse to the direction of propagation (gauss), and H_L is the component along the direction of propagation. Let x be defined such that

$$n^2 = 1 - \frac{kN}{x}. \tag{8}$$

If the magnetic field is neglected, then x is the square of the frequency. However, the effect of the magnetic field can be partially corrected for by choosing some "average" value for the field in that region of the ionosphere traversed by the ray, and then using this to compute the two values of x corresponding to the "O" and the "X" modes. From (4), it is seen that only the magnitude and not the sign of the magnetic field components is important in determining an average. At the higher frequencies of oblique propagation, the magnetic field has less effect on the refractive index than at the lower frequencies of vertical propagation, and so it is a reasonable approximation to include the magnetic field in constructing the electron density distribution and to neglect it in (8) for oblique propagation.

Using Fig. 1, it is seen that

$$\sin \alpha = \frac{rd\theta}{\sqrt{r^2 d\theta^2 + dr^2}}, \tag{9}$$

where r and θ are polar coordinates with the origin at the center of the earth.

Combining (2), (3), (8), and (9) gives

$$\theta = \theta_0 + \sin \alpha_0 \int \frac{d\rho}{\rho \sqrt{\left(1 + \frac{b_i}{x}\right)\rho^2 - \frac{a_i}{x}\rho^4 - \sin^2 \alpha_0}}, \tag{10}$$

where θ_0 is a constant of integration. If ρ^2 is set equal to z, (10) becomes, upon integration,

$$\theta = \theta_0 + (1/2) \sin^{-1}\left[\frac{(x + b_i) - \frac{2x}{z}\sin^2 \alpha_0}{\sqrt{(x + b_i)^2 - 4a_i x \sin^2 \alpha_0}}\right]. \tag{11}$$

Eq. (11) gives the variation of θ with z in the ith segment. The integration constant, θ_0, has a different value for each segment. Let the polar axis be chosen to intersect the earth at the transmitter. Between the earth and the bottom of the ionosphere

$$a_i = a_0 = b_i = b_0 = 0 \tag{12}$$

and at the transmitter

$$\theta = 0, \qquad z = 1, \tag{13}$$

so that below the ionosphere

$$\theta_0 = -1/2 \sin^{-1}[1 - 2 \sin^2 \alpha_0]. \tag{14}$$

Simplifying this gives

$$\theta_0 = \alpha_0 - \frac{\pi}{4}, \tag{15}$$

and hence the variation of θ with z is this segment (the "zeroth" segment) is

$$\theta = \alpha_0 - \frac{\pi}{4} + 1/2 \sin^{-1}\left[1 - \frac{2}{z}\sin^2 \alpha_0\right]. \tag{16}$$

Let the limits of the ith segment defined by a_i and b_i occur at $z = z_{i-1}$ and $z = z_i$ (see Fig. 2), and also let $S_{i,k}$

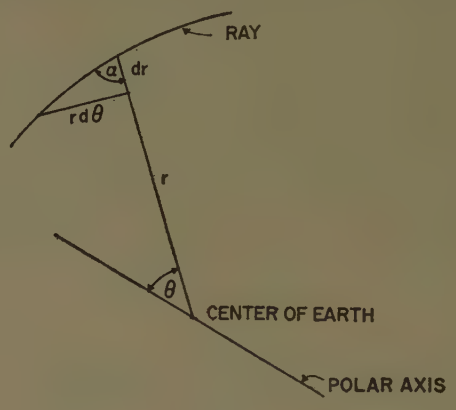


Fig. 1.

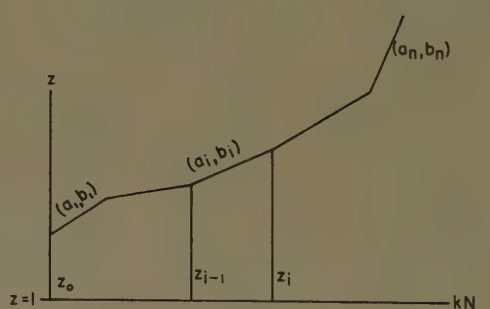


Fig. 2.

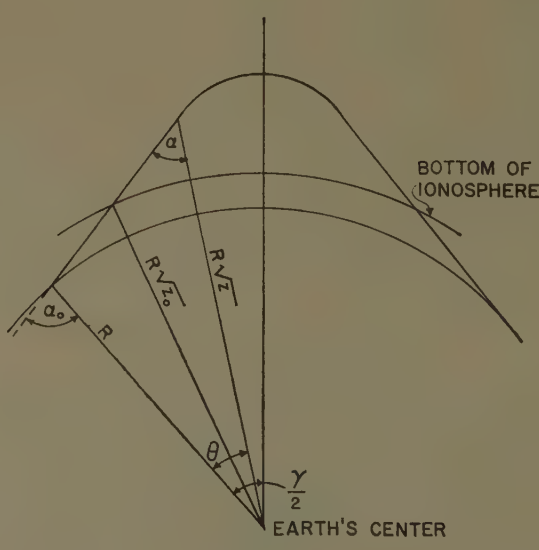


Fig. 3.

represent the following:

$$S_{i,k} = \sin^{-1}\left[\frac{x + b_i - \frac{2x}{z_k}\sin^2\alpha_0}{\sqrt{(x+b_i)^2 - 4a_i x \sin^2\alpha_0}}\right]. \quad (17)$$

For z in the jth segment, that is, for z in the range

$$z_{j-1} < z < z_j,$$

the variation of θ can be determined by repeating the procedure used to obtain (16), for each segment.

$$\theta = \alpha_0 - \frac{\pi}{4} + 1/2 \sum_{i=0}^{j-1} [S_{i,i} - S_{i+1,i}]$$

$$+ (1/2)\sin^{-1}\left[\frac{x + b_j - \frac{2x}{z}\sin^2\alpha_0}{\sqrt{(x+b_j)^2 - 4a_j x \sin^2\alpha_0}}\right]. \quad (18)$$

This is the equation of the ray path, $\theta = \theta(z)$, from the transmitter to the top of the path. At the top of the path where

$$\frac{d\rho}{d\theta} = 0 \quad (19)$$

the surd in (10) is zero. If the top of the path occurs in the jth segment, then z can be eliminated from the last term in (18). The last term becomes $\pi/4$ radians after substitution. The total angle, γ, subtended at the center of the earth by the ray is then

$$\gamma = 2\alpha_0 + \sum_{i=0}^{j-1} [S_{i,i} - S_{i+1,i}]. \quad (20)$$

The next term in the series $S_{j,j}$ is an inverse sine of a number greater than unity. Thus to obtain γ the series is continued until the first unreal term is obtained. If all the segments used to approximate the electron density distribution produced real terms, then either the ray penetrates the ionosphere or there is insufficient data to plot the complete ray.

Application to Ottawa-Slough Path

The above method was applied to the Ottawa-Slough oblique path to study the one hop mode. The oblique ionograms for 1640 GMT and 1700 GMT, October 27, 1957 (see Figs. 4 and 5), were used. These times correspond to local times at the midpoint of the path of about 1400 LT and 1420 LT. These records were used since they appear to have a direct ray on the one hop mode. From the records, it cannot be determined by inspection whether this is a true direct ray or a distortion of the Pedersen ray due to a stratification at high F layer heights.

The ionospheric distribution used was constructed by

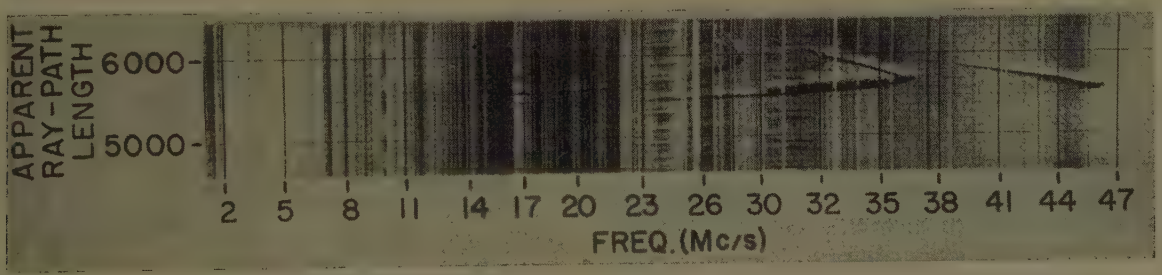


Fig. 4—Ottawa-Slough oblique ionogram for 1640 GMT, received at Ottawa, October 27, 1957.

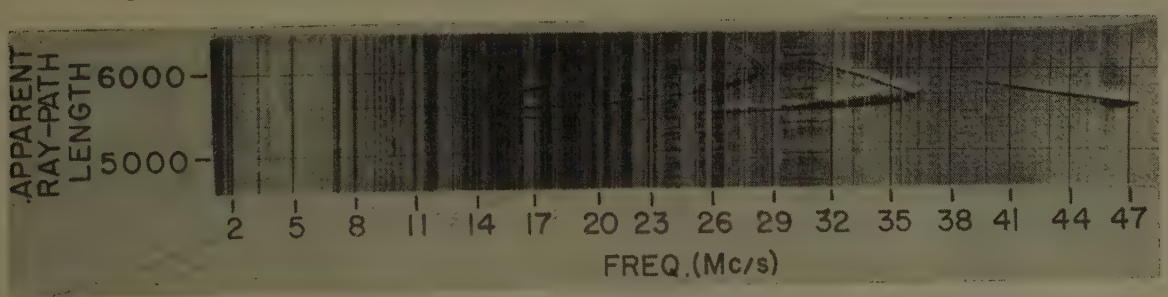


Fig. 5—Ottawa-Slough oblique ionogram for 1700 GMT, received at Ottawa, October 27, 1957.

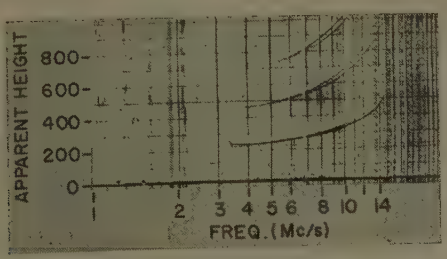


Fig. 6—Vertical ionogram for 1400 AST recorded at St. John's, Newfoundland, October 27, 1957.

Schmerling's method (3), from the vertical ionogram shown in Fig. 6, recorded at St. John's, Newfoundland at 1400 AST (1430 Local Time). The magnetic field at St. John's is very similar to that at Washington, D. C., and since Schmerling had done the preliminary calculations for Washington, these calculations were used to construct the electron distribution at St. John's. The electron density distribution was divided by eye into three straight line segments as shown in Fig. 7. The constants a_i, b_i and z_i were obtained from each segment, and using (8), Figs. 8 and 9 were constructed. In making

$$a_i\rho^2 + b_i = a_{i+1}\rho^2 + b_{i+1} \tag{21}$$

at z_i, it was found that five or six significant figures were necessary for the constants. Fig. 8 shows the variation of the angle γ, subtended at the earth's center, with the angle of departure, α_0, for two fixed frequencies, and Fig. 9 shows the variation of frequency with the angle of departure for the Ottawa-Slough path. It was possible to compute only a small portion of the Pedersen ray in Figs. 8 and 9 due to the disappearance of the vertical ionogram trace near the critical frequency. Fig. 9 shows that earth cutoff is to be expected close to the maximum usable frequency in such a way that a small part of the direct ray is predicted. This is fairly conclusive evidence that the one hop modes shown in Figs. 4 and 5 do contain direct rays.

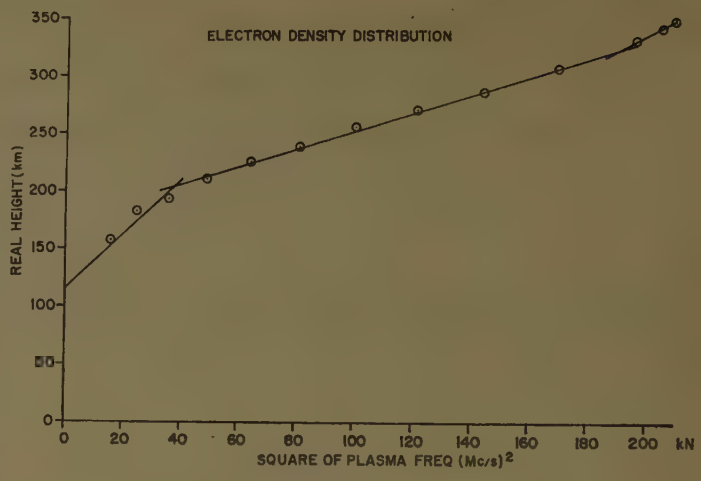


Fig. 7.

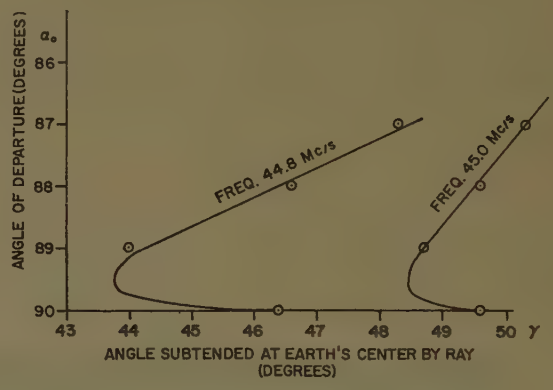


Fig. 8.

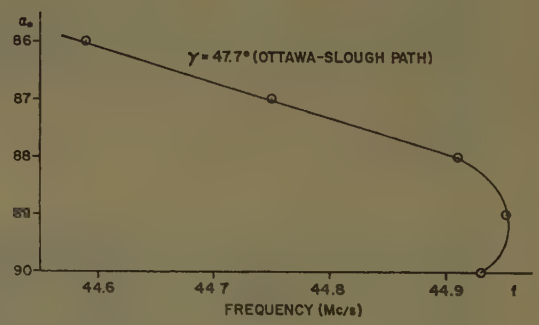


Fig. 9.

Conclusion

The above ray-tracing technique may be applied to many problems in oblique propagation where the conventional methods of analysis would be inaccurate or uncertain. These methods involve either the assumptions that the equivalence theorem holds or that the vertical electron density distribution has some particular shape. The labor involved by this ray tracing technique is roughly proportional to the number of straight line segments used; consequently, good accuracy is relatively easy to obtain. This method seems well suited to the calculation of ray paths by a digital computer.

The Effect of Multipath Distortion on the Choice of Operating Frequencies for High-Frequency Communication Circuits*

D. K. BAILEY†

Summary—**Harmful multipath distortion on high-frequency facsimile services and telegraphic services operating at high speeds occurs when the received signal is composed of two or more components arriving by different modes over the same great-circle path with comparable intensities, but having travel times which differ by an amount equal to an appreciable fraction of the duration of a signal element. The dependence of multipath distortion on the relationship of the operating frequency to the MUF is discussed and a new term, the multipath reduction factor (MRF), is introduced which permits calculation in terms of the MUF of the lowest frequency which can be used to provide a specified measure of protection against multipath distortion. The MRF has a marked path-length dependence and is calculated as a function of path length for representative values of the other parameters involved by making use of an ionospheric model. It is then shown how the MRF can be used in connection with world-wide MUF prediction material to determine the minimum number of frequencies which must be assigned to a high-frequency communication service of continuous availability operating at high speed. Some comparisons with observations are discussed, and finally conclusions are drawn concerning manner of operation and choice of operating frequencies to reduce or to eliminate harmful multipath distortion.**

Introduction

MULTIPATH distortion is a matter of serious concern to such high-frequency communication services as automatic telegraphy and facsimile; it is of less serious concern for telephony. Multipath distortion occurs when the received signal consists of two or more components arriving with comparable intensities over the same great-circle path by different modes and having different times of travel. The ratio of the difference in arrival times of the first and last significant components of a signal to the length of a typical signal element measures the severity of the multipath distortion. The term multipath distortion as applied to high-frequency radio communication will be understood here to exclude the deleterious effects of echo[1] and signals arriving from other than the great-circle direction of the transmitter. In the latter category are found weak signals composed of many components of different delays which arrive by means of the scatter mechanism. Scattered signals are usually too weak to cause difficulty as long as the service is operating on a frequency below the MUF as normally understood. Harmful multipath signal components are seldom encountered with delays much in excess of 4 or 5 milliseconds, whereas echo is characterized by delays from 10 to 100 milliseconds or more.

Relationship to Operating Frequency

The number of modes received over a high-frequency communication path depends on the relationship of the actual operating frequency to the MUF for the path. If the operating frequency is equal to or just below the MUF, it is usually possible to receive by means of a single mode, and multipath distortion is entirely absent. As the operating frequency is decreased relative to the MUF, additional modes of transmission become possible, corresponding to the increased number of separate reflections between the reflecting layer and the earth, each consisting of shorter individual "hops" and having, therefore, lower MUF's. The number of modes received also depends upon other factors such as antenna characteristics at the terminals and the attenuation characteristics of the path. In general, higher-order modes—order is here specified as the number of ionospheric reflections involved—are more heavily attenuated, particularly by day when both ionospheric attenuation and ground reflection losses are combined. It is quite possible, however, for the lowest order or first possible mode to be received more weakly than the second or other higher-order modes if the former comprises one or a series of reflections, each having lengths very close to the upper limit which is geometrically possible for a single reflection as determined by the reflection height and curvature of the earth. Under such circumstances the vertical angle of departure and arrival of the signal is so small that the antennas used may be relatively inefficient as compared with their efficiency at higher angles appropriate to higher-order modes.

Considering, first, paths several thousand kilometers in length, it is well known that frequencies far below the MUF cannot be used during daylight, owing to the considerable ionospheric attenuation they undergo and the consequent extremely low signal-to-noise ratio. For such paths the day frequency is usually sufficiently high and multipath distortion becomes most serious when the evening transition condition develops. A day frequency often remains useful until night has fallen on the path, and the normal ionospheric absorption has decreased virtually to zero. When the day frequency finally fails owing to skip, it is found that much lower frequencies

* Manuscript received by the PGAP, June 17, 1958. This paper was originally issued as a Natl. Bur. of Standards Rept., Washington, D. C., 1951. The material in this paper constitutes one of the necessary preliminary investigations undertaken by the author in connectiob with his duties as Chairman of the Propagation Working Group of the Provisional Frequency Board of the International Telecommunications Union, which met at Geneva, Switzerland, from 1948 to 1950. It reflects the international character of the Geneva activities. The Propagation Working Group was specifically charged with the duty of establishing, on an engineering basis, rules for the selection of operating frequencies for high-frequency point-to-point radio communications.

† Page Commun. Eng., Inc., Washington, D. C. Formerly with Natl. Bureau of Standards, Washington, D. C.

[1] D. K. Bailey, "The effect of echo on the operation of high-frequency communication circuits," IRE TRANS. ON ANTENNAS AND PROPAGATION, vol. AP-6, pp. 325–329; October, 1958.

are receivable with satisfactory signal-to-noise characteristics and there is a real danger that too drastic a shift downward to a night frequency will have the result of inflicting intolerable multipath distortion on the service. When this happens it is necessary, in the absence of a suitable intermediate frequency, to reduce the speed of transmission of information for a period of time varying from several minutes to several hours depending on the rate of fall of the MUF for the path. Clearly the ratio of the night frequency to the day frequency—at the time of the frequency change the day frequency and the MUF are presumed equal—is an important measure of the susceptibility of the service to multipath distortion.

Turning now to much shorter paths, a few hundred kilometers in length, it is well known that a number of modes besides the first possible, now identical with the first-order reflection, are present simultaneously even on frequencies only a little below the MUF. This happens because the path lengths are no longer very great by comparison with the reflection height, with the consequence that the MUF for a second-order mode, for example, is not greatly less than that for a first-order mode. Furthermore, the relative attenuation during daylight hours of a second-order mode compared with the first-order mode is not so great; as a consequence multipath distortion may be a serious problem at all hours. The situation is further complicated in detail by the possibilities of significant modes of propagation by layers lower than the F2 layer. For reasons which will become evident when the figures are discussed, this latter complexity need not cause concern as far as the practical side of multipath protection is concerned, since regardless of the precise modes which give rise to a particular case of multipath distortion, high-speed services of the type under discussion cannot be successfully operated on a regular and continuous basis by means of normal ionospheric propagation over distances less than several hundred kilometers. The practical difficulties of providing, by sufficient choice of frequencies and by flexibility in frequency changing, a truly adequate protection against multipath distortion for high-speed telegraphy or facsimile services operating over distances of a few hundred kilometers are apparent and require little further discussion.

Multipath Reduction Factor—Concept and Definition

It will be evident by now that the severity of multipath distortion is intimately related to the ratio of the operating frequency to the MUF for the path; since these are identical during evening frequency changes, it must also be directly related to the ratio between adjacent frequencies in the group or complement of frequencies available to a particular service. In order to proceed to a quantitative discussion of the problem of protecting high-speed services against multipath distortion through correct choice of operating frequencies, it is convenient to introduce and compute a new quantity called the multipath reduction factor,[2,3] abbreviated "MRF." The multipath reduction factor is a different and quantitative way of regarding an otherwise well-known situation; it is specifically intended to be useful in problems of frequency assignment and utilization.

The multipath reduction factor at a given moment for a given path is defined as the ratio of the lowest frequency which will provide a specified degree of protection to the maximum usable frequency. The degree of protection provided is expressed in milliseconds and represents the maximum tolerable delay between the first and last received components of a signal element (*i.e.*, the maximum tolerable elongation of a signal element) consistent with a particular type and quality of service. The determination of the lowest frequency which will provide a specified degree of protection requires the determination of the highest-order mode having a delay relative to the first possible mode equal to or just less than the required degree of protection, and then the computation of the MUF corresponding to the next higher-order mode (the mode which must be excluded at the receiver). This MUF is the frequency required and its ratio to the MUF for the path is the multipath reduction factor.

It can now be seen that the MRF is a complicated function of the characteristics of the ionosphere over the path. For the MRF to be amenable to calculation, some simplifying assumptions are necessary. Of these the assumption that the reflections are essentially specular and take place from a constant height corresponding to propagation via the F layer is the most important. This makes possible relatively simpler calculations of ray length and time of travel assuming the velocity to be that of light. Because the actual equivalent heights of reflection are dependent on the transmission distance and the detailed characteristics of the ionosphere at each reflection point, it is impractical to make rigorous calculations of time of travel. Under these conditions the reduction factor for distances less than 4000 km in terms of the standard MUF's for 4000-km paths is a simple function only of transmission distance for a specified reflection height, and can be directly determined from a conventional transmission curve.[4] The transmission curve shown in Fig. 6.6 of Ref. 4 has been used in the calculations leading to the curves presented in Figs. 1 through 4. The only precaution necessary in the calculations is that of calculating the ray lengths and travel times as a function of distance for the 1F mode with sufficient precision to permit the calculation of the times of travel of higher-

[2] International Telecommunications Union, Provisional Frequency Board, "Second Report of Working Group 1 (Propagation), Committee 4," Doc. No. 375–E, Geneva, Switzerland, p. 3; October 5, 1948.

[3] International Telecommunications Union, Provisional Frequency Board, "Sixth Report of Working Group 1 (Propagation), Committee 4," Annex to Doc. No. 557-E, Geneva, Switzerland, pp. 1–2; April, 1949.

[4] "Ionospheric Radio Propagation," National Bur. Standards, Washington, D. C., Circular 462, pp. 72, 74–76, 101; June 25, 1948.

order modes without significant accumulation of error by simple multiplication of the results of the basic calculation. It is the differences in travel times as a function of distance between the modes governing the particular multipath protection which must be determined accurately.

Multipath Reduction Factor—Illustrative Calculations

For many years it has been recognized that the results of operation of high-frequency circuits cannot be interpreted solely in terms of a strict geometrical ray treatment of the propagation. Nevertheless ray treatments are the only useful means available for attempting to understand many experimentally obtained results. Direct observations, for example, of MUF are in fairly satisfactory agreement with theory for propagation over fairly short distances for which the 1F mode is strongly present. For greater distances the observed MUF's are usually higher than can be predicted from vertical-incidence observations by considering possible modes and determining the lowest MUF for the series of reflection points for the mode of lowest possible order. Early recognition of this discrepancy gave rise to a method of MUF calculation for all circuits having lengths in excess of 4000 km known as the two control-point method.[4] This empirical procedure results in MUF's which are often higher than can be calculated by the more detailed mode treatment; such MUF's are never lower than mode-determined MUF's—nevertheless, they too are often lower than the MUF's required by the observed performance of communication circuits. Despite these difficulties, at present under investigation, pulse experiments between England and New Zealand[5] can be fairly intelligibly interpreted by means of an elementary mode analysis. This success for mode analysis provides a basis for some confidence in the results for long distances presented in Figs. 1 through 5.

Fig. 1 represents the MRF as a function of path length. The reflection height has been taken as 300 km corresponding to F-layer reflection. The results are for a multipath protection of 2 milliseconds, and the results for distances greater than 4000 km are referred to a path MUF determined by the two control-point method, and therefore to a 4000-km MUF. It has further been assumed that the only limits on useful propagation of a particular mode are layer height and earth curvature, so that a particular mode is considered as fully present until its angle of departure reaches zero degrees (until it becomes a tangential ray). The calculations have been extended to a path length of 30,000 km corresponding to long-path operation on a circuit having a short-path length of 10,000 km. The values published at Geneva, and adopted for use by the Provisional Frequency Board,[2,3] are plotted for comparison. They were based on fuller results of the kind shown here. The curve is made up of a number of discontinuous segments which are labelled to show the order of the first mode which must be excluded in order to provide the specified protection. An indication is also given of the first possible mode for all path lengths.

Fig. 2 represents the MRF for identical protection and layer-height values as for Fig. 1, but with the difference that beyond a path length of 4000 km the MRF's are calculated by reference to the MUF for a single "hop" of the first possible mode at each distance. The results below 4000 km are naturally identical to those in Fig. 1. Since in any practical situation the height of reflection is far from constant along the path, and is a function of the actual length of a "hop," a smooth curve has been drawn to represent the data in a manner more suited to practical use, and intercomparison.

Fig. 3 illustrates, for the same height of reflection, the dependence of the MRF on the degree of protection to be provided. It will be seen that operating frequencies must be changed downward in the evening in smaller steps to provide protection for the higher-speed services having smaller maximum tolerable delays. These curves are smoothed in accordance with the discussion of Fig. 2 above.

Fig. 4 illustrates, on the same scale as the preceding figures, the dependence of the MRF on the height of reflection for a 2-millisecond protection. Three heights are shown: 200, 300, and 400 km. It will be seen that larger frequency changes can be safely made with lower layer heights. It is this fact which comes to the partial rescue of communication circuits of a few hundred kilometers length during daytime when circumstances are favorable for fairly high E-layer MUF's and consequently for predominant E-layer modes of propagation.

Fig. 5 is designed to illustrate for the standard case adopted as a guide by the Provisional Frequency Board the range of vertical angles as a function of path length within which the components of the received signal must be kept if multipath protection is to be provided. The lower curve in each case represents the vertical angle corresponding to the first possible mode, and the upper curve of the vertical angle corresponding to the first mode which must be excluded. This figure is intended to be of assistance in problems of antenna design. In this connection several further calculations have been performed to illustrate the effect on the MRF of assuming that no useful modes can exist which have vertical angles of radiation less than 2° and 5° above the horizon. These assumptions are intended to represent situations which can occur with some practical types of antennas. They result in discontinuous curves similar to those of Figs. 1 and 2, but differing as to detail. The significant point is that the smooth curve through them does not sensibly differ from that shown in Fig. 2 for the same height of reflection and protec-

[5] "New Zealand-United Kingdom Radio Circuit, Report on Propagational Effects," New Zealand Post Office Engineering Branch Res. Rep. No. 2, General Post Office, Wellington, N. Z.; June 15, 1945. (Private communication.)

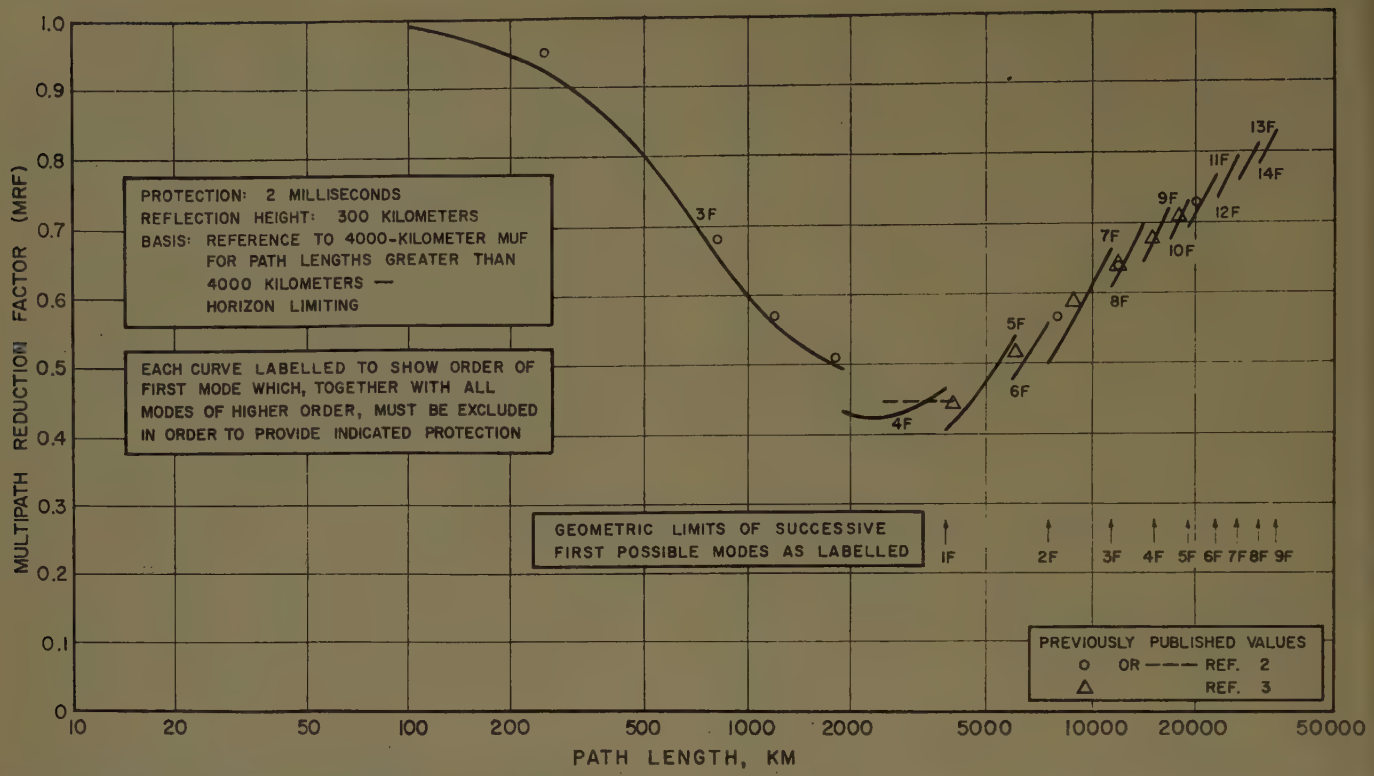


Fig. 1—Multipath reduction factor (MRF) for 2-millisecond protection as a function of path length, for a fixed reflection height of 300 km showing detailed values referred to 4000-km MUF's for distances greater than 4000 km, and showing previously published values.

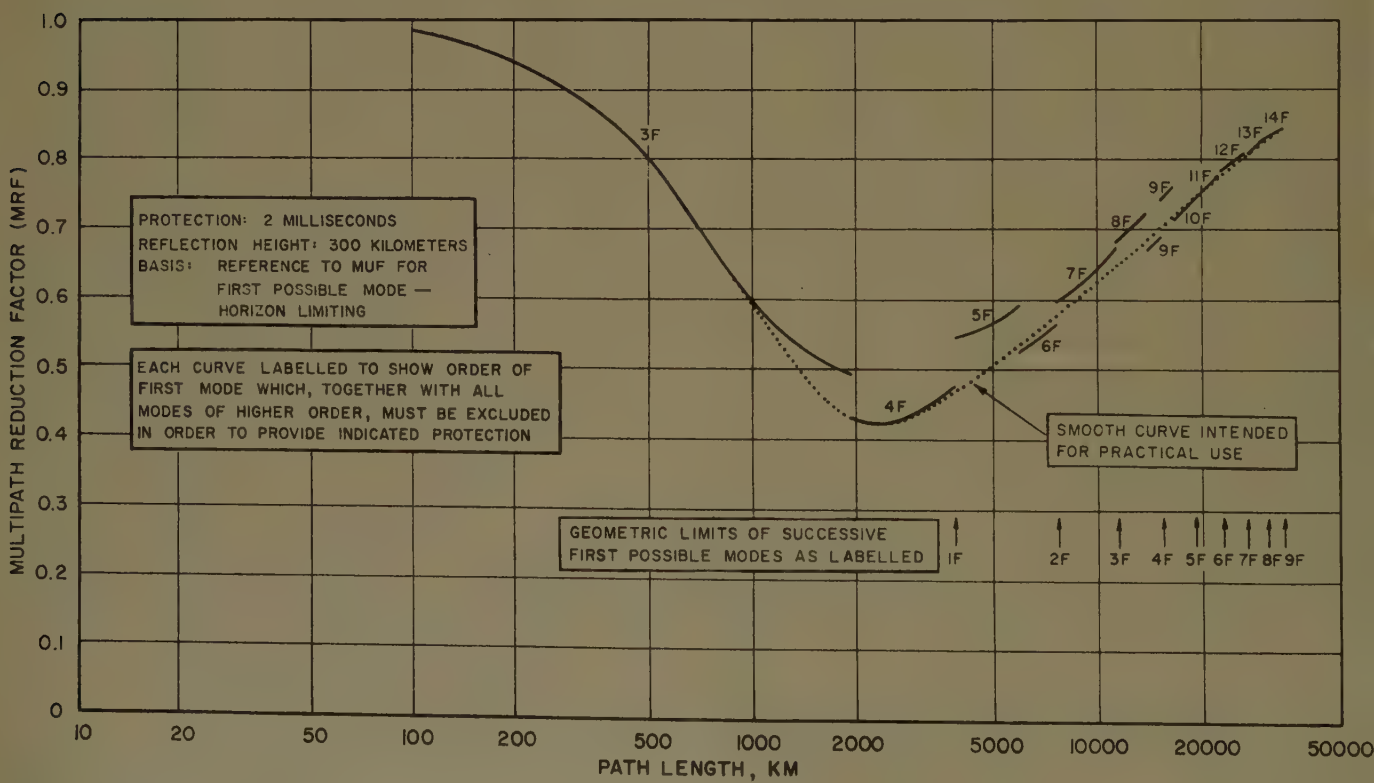


Fig. 2—Multipath reduction factor (MRF) for 2-millisecond protection as a function of path length, for a fixed reflection height of 300 km, showing detailed values referred to the MUF for the first possible mode for distances greater than 4000 km, and showing the smooth curve intended for practical use.

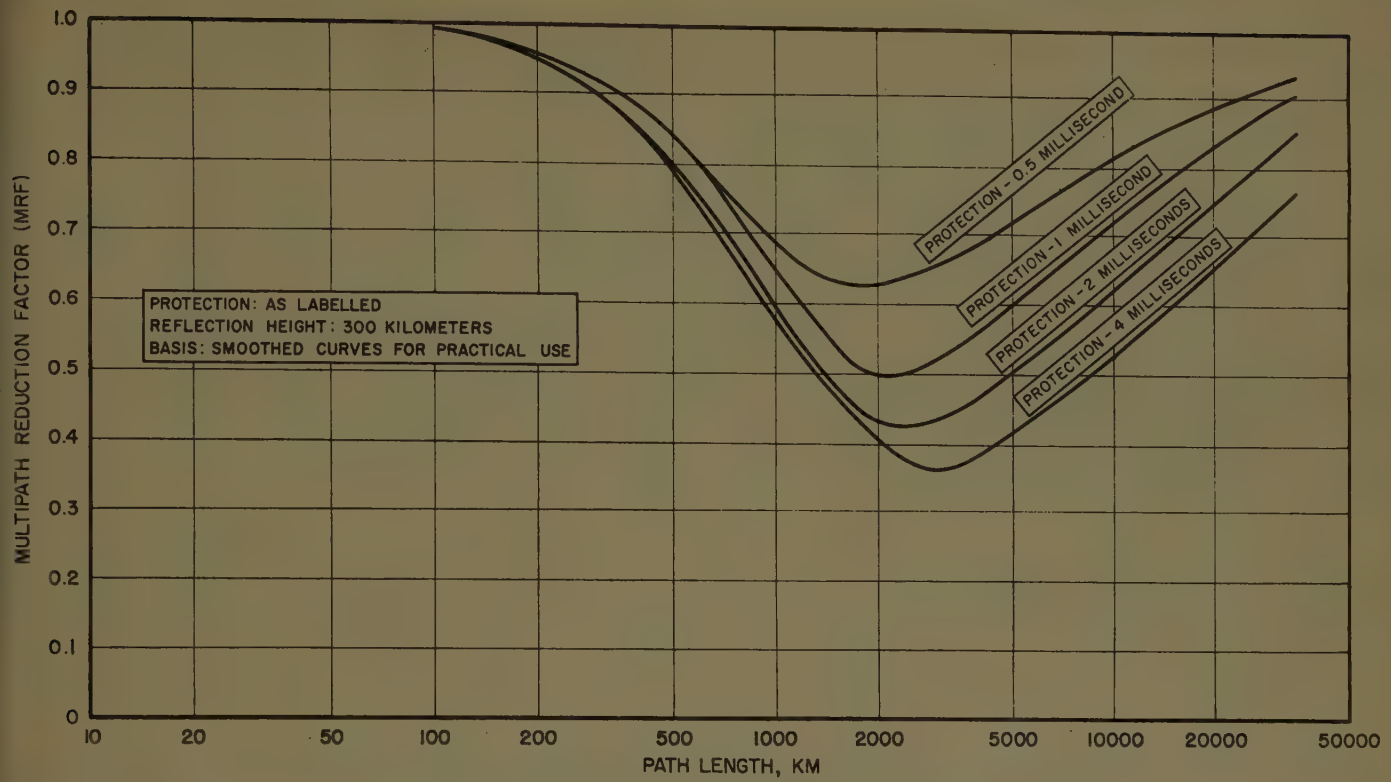


Fig. 3—Smoothed curves of multipath reduction factor (MRF) for a fixed reflection height of 300 km, as a function of path length illustrating the dependence on desired protection expressed in milliseconds.

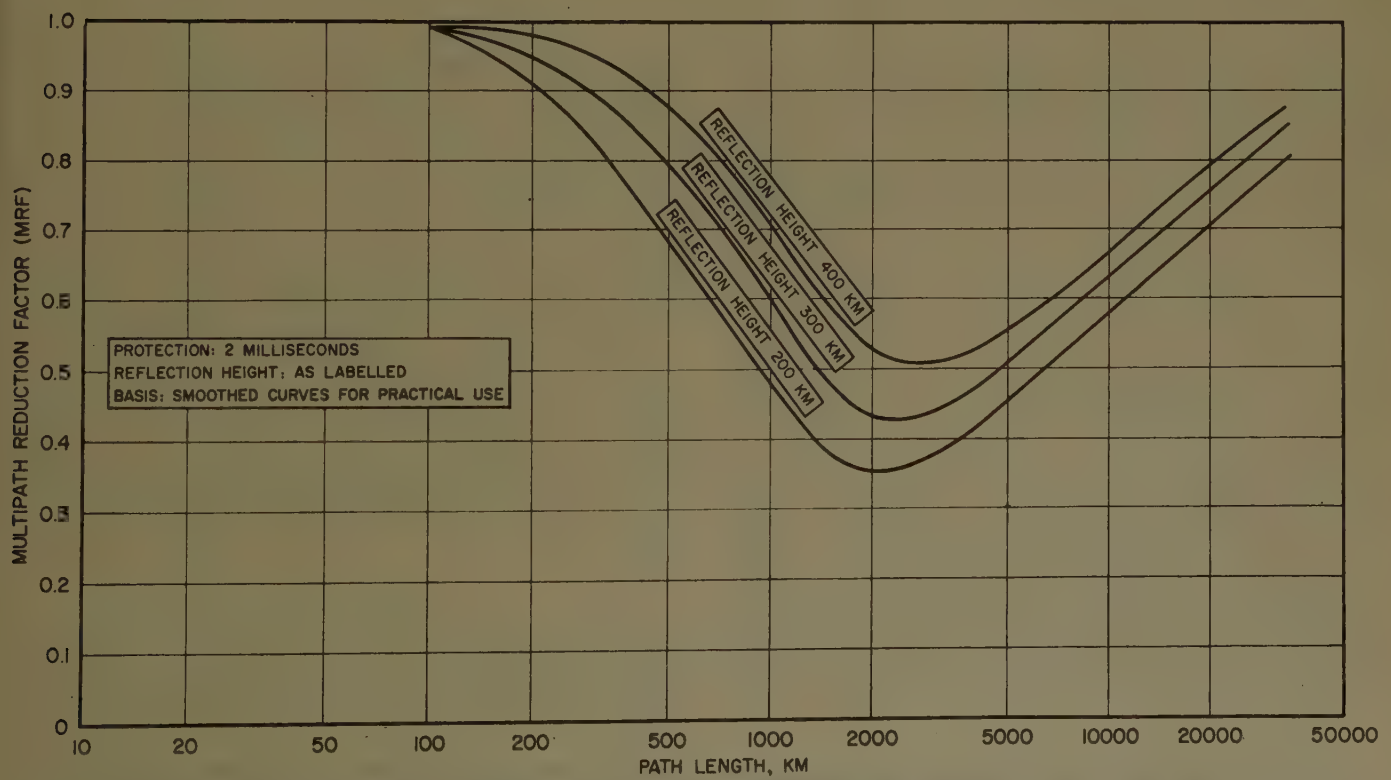


Fig. 4—Smoothed curves of multipath reduction factor (MRF) for 2-millisecond protection as a function of path length illustrating the dependence on the value of fixed reflection height assumed.

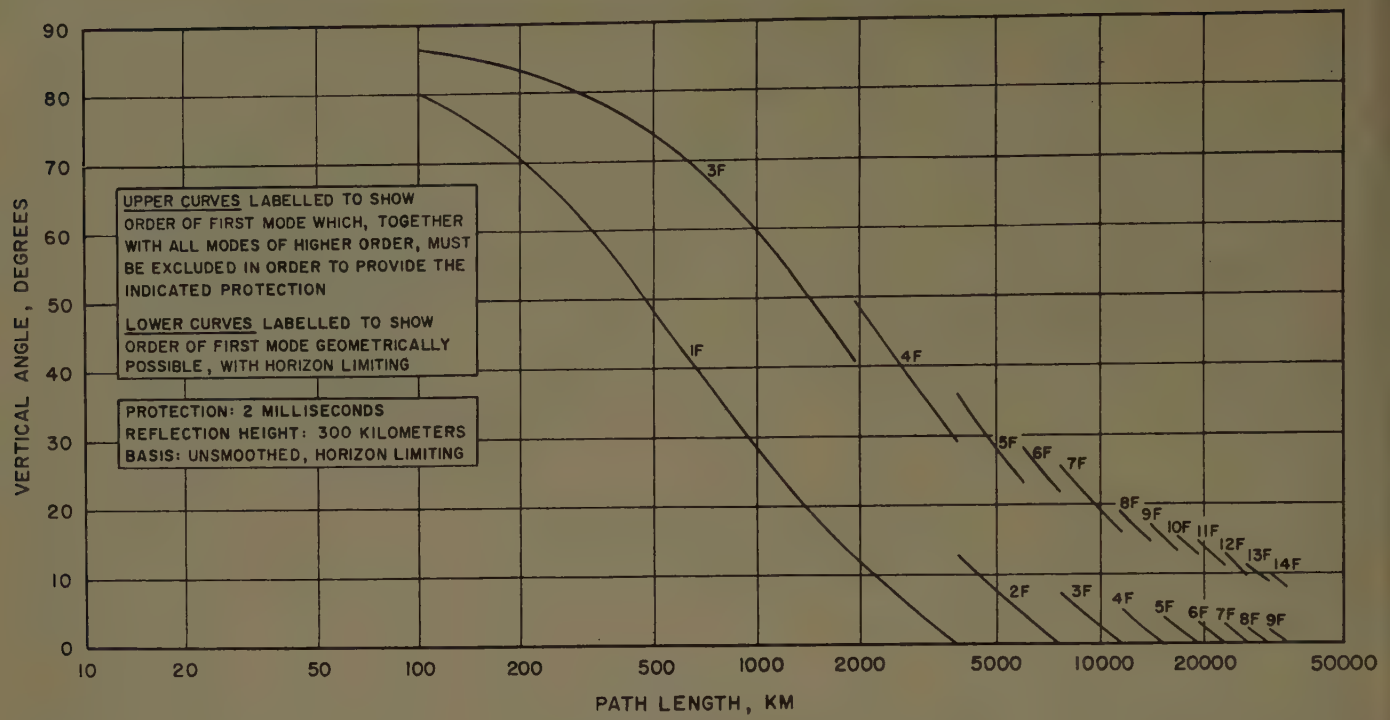


Fig. 5—The range in vertical angle of arrival as a function of path length within the components of the received signal, corresponding to different modes of propagation, must be kept in order to provide 2-millisecond protection for a fixed reflection height of 300 km.

tion value. For vertical-angle limitations close to but above the horizon, Fig. 5 would be changed only in details of the discontinuities. If smooth curves were drawn in Fig. 5 they would be raised approximately by the amount of the angle limitation, as long as the limitation was not much over 5°.

Application of the MRF to Frequency Assignment Problems

In the matter of assigning sets or complements of operating frequencies to high-speed communication circuits, the usefulness of the MRF can best be understood with the aid of two auxiliary quantities, now defined. Let the max-max-MUF[6] represent the highest monthly median MUF, against the occurrence of which it is desired to provide multipath protection. This value occurs at a particular hour of a particular month, usually in the year of highest solar activity. Similarly let the min-min-MUF[6] represent the lowest monthly median MUF. This value occurs at a particular hour of a particular month, a winter month in temperate regions, in the year of lowest solar activity. The lowest operating frequency assigned to a service must be low enough to permit operation under normal conditions at times when min-min-MUF conditions prevail. Both the max-max-MUF and the min-min-MUF can be predicted from available material.[6] The MRF now provides a tool for fixing a lower limit to the number of operating frequencies required if protection against harmful multipath distortion is to be provided at all times. Represent the MRF for the particular protection and path length as MRF (p, d) and the required operating frequencies in descending order by $f_1, f_2, \cdots, f_n$, where n is the number required and f_n is either equal to or determined in relation to the min-min-MUF. Then it is possible to write

$$\frac{f_1}{\text{max-max-MUF}} \geq \frac{f_2}{f_1} \geq \frac{f_3}{f_2} \geq \cdots \geq \frac{f_n}{f_{n-1}} \geq \text{MRF } (p,d)$$

where n is chosen to be the lowest integer which will satisfy the relations (*i.e.*, when equality most nearly satisfies the relations). It should, of course, be understood that other engineering considerations are involved in frequency assignments which may make it necessary to

[6] "High Frequency Radio Propagation Charts for Sunspot Minimum and Sunspot Maximum," Rep. CRPL-1-2, 3-1, Central Radio Propagation Lab., National Bur. Standards, Washington, D. C.; December, 1947.

assign more frequencies. For very short paths, over which high-speed working should not ordinarily be attempted on a continuous basis, fewer high frequencies are customarily assigned.

Comparison with Observations

The comparison with observations is a rather difficult matter. The results of the England-to-New Zealand pulse experiments have already been mentioned. Examples of destructive multipath distortion on the service from London to Capetown have been examined,[7] where the night frequency tried was about 0.50 times the day frequency. The path is about 9600 km in length and the maximum reduction permissible according to the curve of Fig. 2 is by a factor of 0.62. On the whole, Eckersley, in two classic papers,[8,9] has supplied the most beautiful early evidence in support of the approach to multipath distortion made in this paper. While it is virtually impossible to extract actual values for the MRF from Eckersley's data, it is not difficult to appreciate that his data are nowhere in conflict with the present findings. Hallborg[10] has attempted a fairly detailed study of the New York-to-San Francisco path including considerations of multipath distortion.

Engineering opinion varies as to the safe upper limit for multipath elongation of a signal element, a most understandable situation when it is borne in mind that matters such as fading, and the effects of widely varying received intensities of the multipath components, though clearly involved, have not been considered. Engineers of Cable and Wireless Ltd. report that they like to keep the elongation to 20 per cent or less, whereas RCAC operation can apparently stand elongations as high as 40 to 50 per cent.[11] The Provisional Frequency Board was content to accept a 2-millisecond protection as adequate for high-speed services requiring multipath protection. This standard was adopted for guidance as part of the engineering basis for the selection of complements of frequencies prior to attempting to make detailed frequency assignment plans. The situation for facsimile was not investigated in detail, and it must be left to users to establish what protection they require. In any event, with facsimile it is usually possible to choose limited periods of operation sufficiently critically to obtain good results, particularly on the highest frequencies available to the service.

[7] A. M. Humby, private communication; 1948.
[8] T. L. Eckersley, "Multiple-signals in short-wave transmission," Proc. IRE, vol. 18, pp. 106–122; January, 1930.
[9] T. L. Eckersley, "Studies in radio transmission," *J. IEE*, vol. 71, pp. 405–459; September, 1932.
[10] H. E. Hallborg and S. Goldman, "Radiation angle variations from ionosphere measurements," *RCA Rev.*, vol. 8, pp. 342–351; June, 1947.
[11] "RCA Comments Regarding Multi-Path Effects in the Operation of Radio Circuits," United States Delegation, Provisional Frequency Board, Geneva, Switzerland, Tech/10; October 11, 1948. (Private communication.)

Summary and Conclusions

The following general conclusions about multipath distortion of high-speed services may be stated; some of them are well known, but require examination against the background of the present approach via the multipath reduction factor.

1) The harmful effects of multipath distortion cannot in a particular situation be reduced by increasing the power.

2) The harmful effects of multipath distortion can always be reduced or eliminated by increasing the operating frequency, or by reducing the speed of transmission of information. Fig. 5 provides information which suggests that there is some possibility of reducing the harmful effects of multipath distortion through careful antenna design.

3) The degree to which a high-speed circuit is subject to multipath distortion is a function of path length. Though this function is discontinuous, it can conveniently be represented as a smooth function. Very short and very long paths are more subject to multipath distortion than paths of intermediate length. Paths of about 2500 kilometers appear to be least affected, though the effects are not severe between about 1500 and 6000 kilometers.

4) The multipath reduction factor (MRF) is a measure of the degree to which a high-speed circuit is subject to multipath distortion; it represents the ratio of the lowest frequency which may be used with satisfactory results (insofar as considerations other than multipath distortion do not intervene) to the MUF for the path at a given moment. As such it is of importance in the selection of operating frequencies, both as one of the considerations involved in the choice of the highest frequency to be assigned, and in the spacing of adjacent frequencies of a complement.

5) The MRF has been calculated as a function of path length for three models of the ionosphere, and for four values of the required protection expressed as the maximum tolerable delay in milliseconds. In general, it will be seen that lower frequencies can be used at a particular moment, according as the maximum tolerable delay is larger, or the reflection height in the ionosphere is lower.

6) The MRF is insensitive to practical limitations which may exist in the efficiencies of antennas at very low angles of departure and arrival.

7) The recommendation made at the International Radio Conference of Atlantic City in 1947,[12] that the "maximum frequency to be used should be taken as approximately 70 per cent of the highest MUF," fails to

[12] "Final Acts of the International Telecommunication and Radio Conferences, Recommendations and Resolutions," Atlantic City, N. J., pp. 75–76; 1947.

recognize the path-length dependence of the MRF and is potentially wasteful of spectrum space. Corresponding to an MRF of 0.70, this recommendation would according to Fig. 2 overprotect, through the assignment of oversized frequency complements, all high-speed services having path lengths, for 2-millisecond protection, between 700 and 15,000 kilometers.

Acknowledgment

The author acknowledges the very helpful counsel and computing assistance received at Geneva from A. M. Humby, formerly of Cable and Wireless Ltd., a member of the United Kingdom Delegation representing the British Admiralty Signals Department, without whose insistence and vast fund of experience this quantitative investigation would not have been possible. Acknowledgment is also due to D. McDonald, Chief of the Australian Delegation, G. E. F. Damant, Chief of the South African Delegation, and R. C. Kirby of the United States Delegation, for very useful personal assistance in the computing. For helpful discussion and results of operational experience, against which the theoretical calculations were checked, thanks are due to Dr. H. C. A. van Duuren, Chief of the Netherlands Delegation, G. Searle, Chief of the New Zealand Delegation, R. Keen, member of the United Kingdom Delegation representing Cable and Wireless Ltd., and C. E. Pfautz, a member of the United States Delegation representing the Radio Corporation of America. In connection with a recent detailed recalculation of the figures particular thanks are due to Miss Loris Perry.

Analysis of 3-cm Radio Height-Gain Curves Taken Over Rough Terrain*

H. T. TOMLINSON† AND A. W. STRAITON‡

Summary—**This report describes the effect of terrain and meteorological conditions on the height-gain pattern of 3.2-cm radio waves over various short transmission paths. Equivalent reflection coefficients are obtained and potential reflection areas are investigated. A study of the time variations in the height of nulls in the signal strength pattern is made and the relationship between movement of the nulls and the corresponding refractive index distribution is considered.**

Introduction

THIS paper describes the analysis of height-gain curves taken over seven line-of-sight paths in the vicinity of Austin, Tex. These height-gain curves were used to obtain equivalent reflection coefficients and attempts were made to identify their principal features with the associated ground profiles. For one of the paths used, a study was made of the variations of the height of several nulls as a function of time.

The emphasis in this paper will be on the data and their analysis with the details of the measuring systems omitted. It is felt that sufficient precautions were taken to avoid instrumental errors that would invalidate the data.

Receiver Site

For each transmission path considered, the receiving site was the 275-foot tower of the Electrical Engineering Research Laboratory at the Balcones Research Center near the north edge of Austin. A view of this tower is shown in Fig. 1 and a view of the receiving equipment mounted in the elevator on the side of the tower is shown in Fig. 2. The height-gain curves were taken as the elevator moved continuously up the tower. The time interval for these measurements was about 2½ minutes. The elevation at the base of the tower is 768 feet MSL.

In additional to the radio data, wet- and dry-bulb temperatures were measured at a number of heights with a psychrometer essentially simultaneously with the radio data.

The Transmission Path

Seven transmitter sites were used in conjunction with the common receiver site. The locations used for the

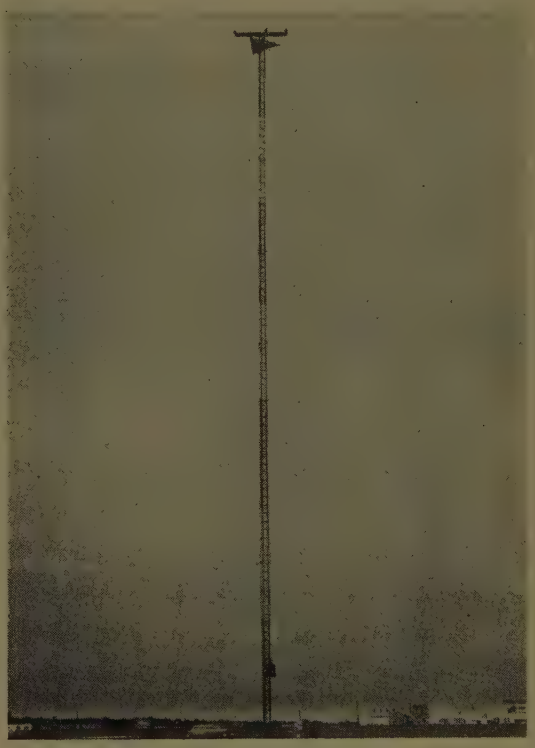


Fig. 1—Receiving site tower.

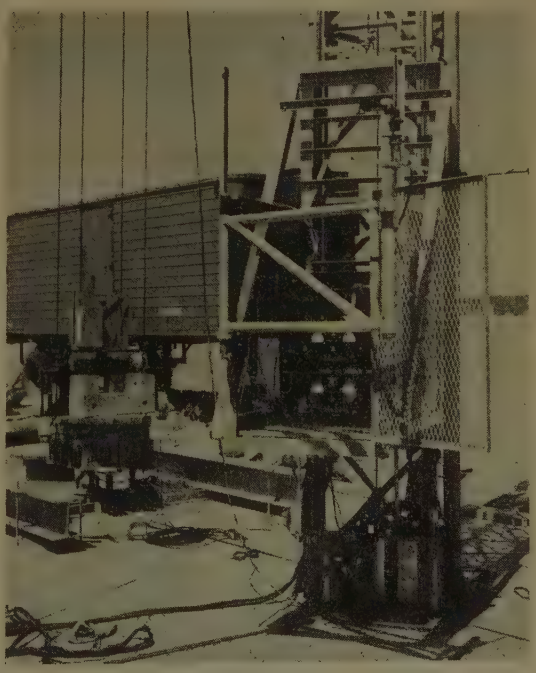


Fig. 2—Receiver elevator equipment.

* Manuscript received by the PGAP, August 4, 1958.
† Chance-Vought Aircraft Corp., Arlington, Tex., formerly at University of Texas, Austin, Tex.
‡ University of Texas, Austin, Tex.

transmitter are shown in Fig. 3 with the designations which will be used later in this paper. Table I indicates the range, transmitter elevation, and general terrain features for each of the paths.

More details concerning each of these paths will be given in later sections when an analysis is made of the resulting height-gain curves.

Height-Gain Translating System

The original recordings of the height-gain patterns, which were made on single-channel Sanborn paper, were on the order of 30 inches long and 2 inches wide. Because these dimensions were obviously incompatible with the normal $8\frac{1}{2}\times 11$-inch page size, a translating system was devised. This system was composed of two main parts; a playback unit and a recording unit.

The playback unit was a Brush recorder movement. The spool guides were brought together sufficiently to hold the narrow Sanborn paper in alignment, and potentiometers were geared to the take-up spool and to a pointer device used to follow the trace on the paper. With low dc potentials applied to the terminals of both potentiometers, voltages were available from the wipers of these potentiometers which were dependent on both altitude and field intensity of the height-gain pattern as represented by distance along the recording and movement of the trace across the paper.

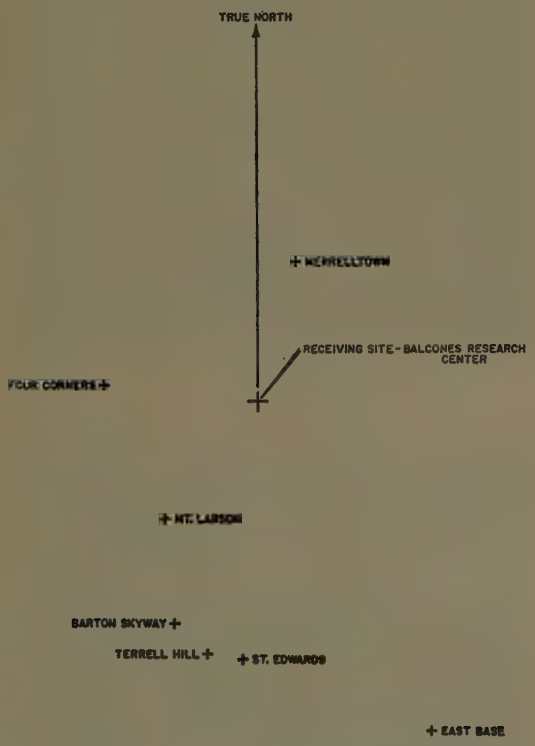


Fig. 3—Transmitter site locations.

TABLE I

Site	Elevation MSL, Feet	Range, Feet	Topographical Features
St. Edwards	680	53,500	Over city
Terrell Hill	680	51,700	Over city
Barton Skyway	690	49,300	Hilly part of city
Mt. Larson	920	33,600	Predominant ridges
Four Corners	1030	31,900	Wooded ridges
East Base	530	76,000	Flat area near transmitter
Merrelltown	930	30,000	Single rounded hill

The recorder unit was a Mosley X-Y recorder, on which both axes of a graph could be controlled at the same time. The movement of the recording pen along either axis was determined by the magnitude of an input voltage to the servoamplifiers driving the pen movement.

The translation of the original recording was accomplished by running the original recording through the Brush movement and manually following the trace with the pointer. As the paper passed to the take-up spool, two varying dc voltages were fed from the wipers of the potentiometers to the servoamplifiers driving each axis of the X-Y recorder. The pen of the X-Y recorder was positioned according to these voltages and traced out a translated height-gain pattern. The gains of the servoamplifiers were adjusted to reduce the size of the altitude axis and to expand the variations of the field intensity.

The altitude information was obtained from the original recordings by marking 10-foot steps on the Sanborn paper and letting the recording pen touch the paper only at the times when these marks passed under the pointer on the Brush playback unit.

The db scales on the original recordings were used to position the pointer on the playback unit and to mark the field intensity scale on the translated pattern.

A photograph of the translating system is shown in Fig. 4.

Theoretical Height-Gain Curves

Flat Earth Reflection

The development of the equation for the field intensity at the receiver as a function of height for a single flat reflector is given in many texts ([2], for example) and will not be given here. The geometry is shown in

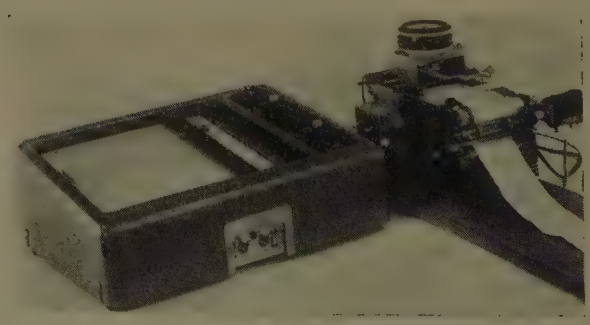


Fig. 4—Height-gain pattern translation system.

Fig. 5 and the phase delay of the reflected wave with reference to the direct is given approximately for near-grazing angles in a homogeneous atmosphere as

$$\Delta\phi = \frac{4\pi z_1 z_2}{\lambda R} + \phi_0 \qquad (1)$$

where

z_1 = height of the transmitter above the reflecting plane
z_2 = height of the receiver above the reflecting plane
λ = wavelength
R = range
ϕ_0 = phase shift on reflection (approximately 180° for near-grazing angles).

The reflected wave appears to come from a source which is the same distance below the reflecting plane as the real transmitter is above. For an area to be a contributing reflecting surface it must fulfill the following conditions:

1) A line drawn from the image source to the receiver must pass through the surface.
2) The angle of incidence and the angle of reflection on the surface must be equal. Stated in another way, the reflecting surface extended must bisect a line from the real to the image transmitter.
3) The surface should be visible from both ends.

Separation of Nulls

In going from one null to the next, the difference in path length between the direct and the reflected rays will be increased by a wavelength and the separation of the nulls, d, is given by

$$d = \frac{R\lambda}{2z_1} \cdot \qquad (2)$$

This is independent of the receiver height for a plane surface extending between the transmitter and receiver. The resulting height-gain would then be as indicated in Fig. 5.

In the measurements described in this report, the height-gain pattern did not show this regularity. In the analysis that follows, the separation of the nulls for a given receiver height is obtained directly from the graphs, and the position of the image transmitter below the actual transmitter z_1' is determined from

$$z_1' = \frac{R\lambda}{d} \cdot \qquad (3)$$

The line drawn from this image transmitter should pass through the effective reflecting area and the plane of this area should bisect the line from the actual to the image antenna.

Movement of Nulls

The change in the height of the nulls for a uniform change in index of refraction is negligible. However, if the average refractive index for the direct path changes relative to the reflected path, the height of a null may change. If this difference in average refractive index between the direct and the reflected ray path is given by Δn and the corresponding change in the height of the null is given by Δz_2, then

$$\Delta z_2 = \frac{R_1^2}{2z_1} \Delta n \qquad (4)$$

where R_1 is the length of the direct ray path.

This relation for the measured data is examined later.

Knife-Edge Diffraction

Another approach which is applicable to the transmission of waves over a ridge is to examine the height-gain curves associated with knife-edge diffraction. The associated geometry is shown in Fig. 6 and the signal strength as a function of the parameter V is given in Fig. 7. Assuming the knife edge to be centrally located

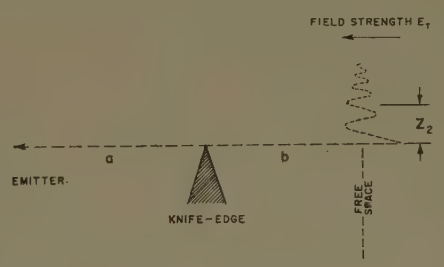


Fig. 6—Illustrations of flat earth reflection and knife-edge diffraction.

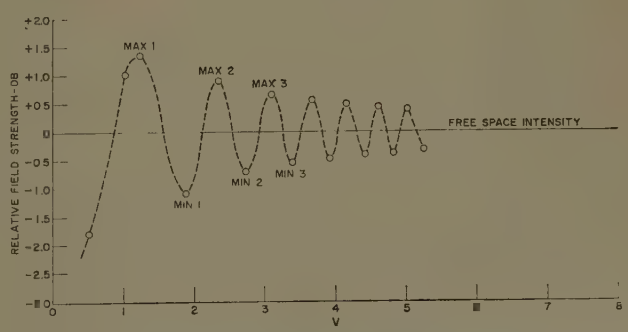


Fig. 7—Distance along cornu spiral (V) vs relative field intensity—db.

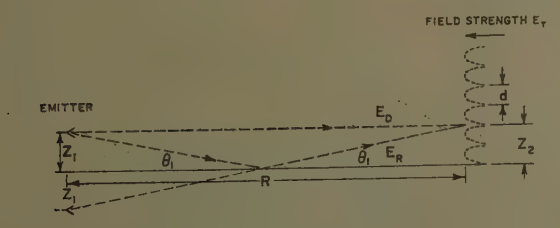


Fig. 5—Flat earth reflection.

on the path, this parameter is related to the distance, Z_2, above the minimum line-of-sight position by

$$V = Z_2/2\lambda D \qquad (5)$$

where D is the horizontal distance from the ridge to the receiver. This height-gain curve has successive lobes decreasing in amplitude and separation as the receiver height is increased. The height-gain function for two of the transmission paths is interpreted in terms of this knife-edge diffraction.

Examination of Measured Height-Gain Curves

Two Examples of Incoherent Height-Gain Curves

Two of the measurement paths designated as St. Edwards and Terrell Hill pass over the business and residential areas of Austin. The profile for the St. Edwards path is shown in Fig. 8 and the height-gain profile is shown in Fig. 9. Similarly, the profile and height-gain curve for the Terrell Hill path are shown in Figs. 10 and 11. All of the major details of the signal strength curves were repeated on successive runs although some of the fine detail varied somewhat.

The height-gain curves for these two paths were very irregular and it is thought that the many buildings along the path which are not indicated on the profiles served to act as partial reflectors. The range from minimum to maximum signal was approximately 3 db for each of the paths. For a single plane wave reflection, this would correspond to a reflection coefficient of about 0.17. Although the single reflection definitely is not applicable, the use of reflection coefficient is a convenient means of describing the range of the fluctuations.

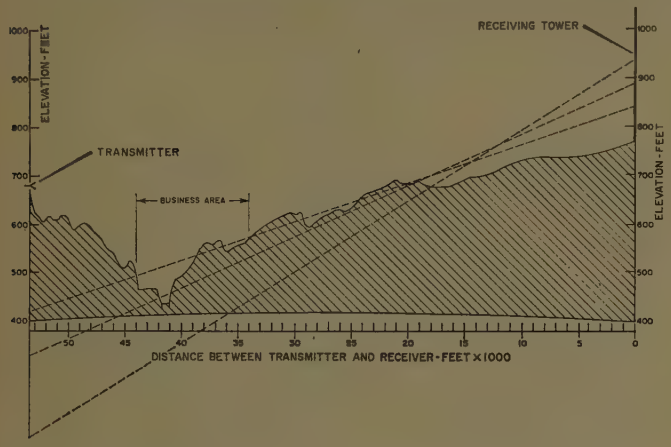


Fig. 8—Elevation profile St. Edwards to Balcones Research Center.

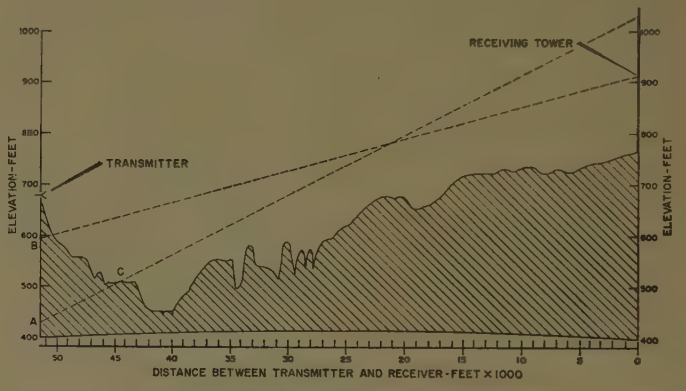


Fig. 10—Elevation profile Terrell Hill to Balcones Research Center.

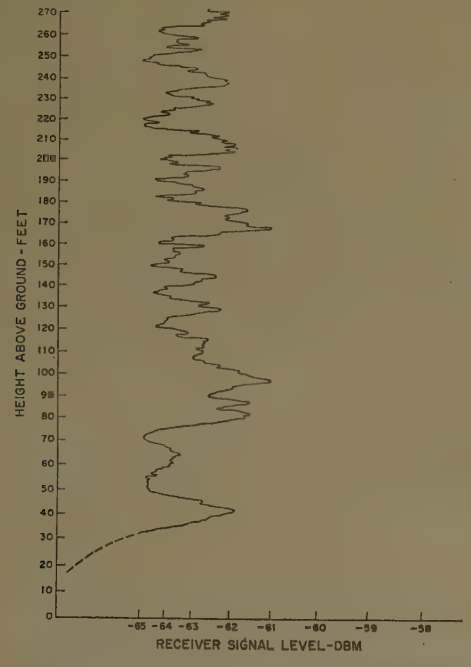


Fig. 9—Height-gain pattern St. Edwards site horizontal polarization. July 18, 1957. 10:50 A.M.

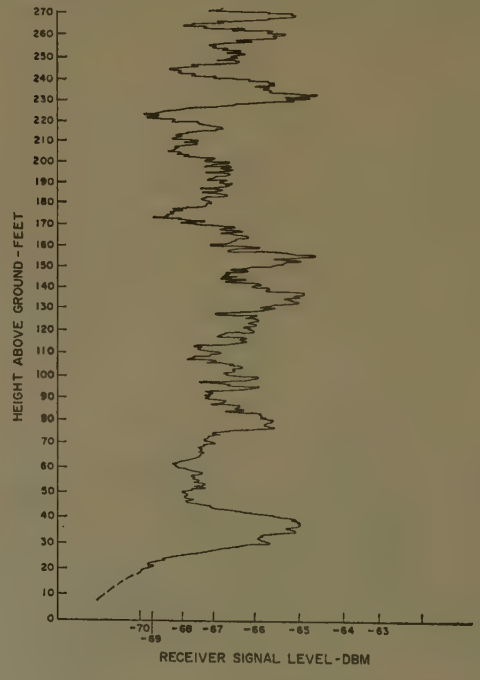


Fig. 11—Height-gain pattern Terrell Hill site horizontal polarization. July 25, 1957. 11:18 A.M.

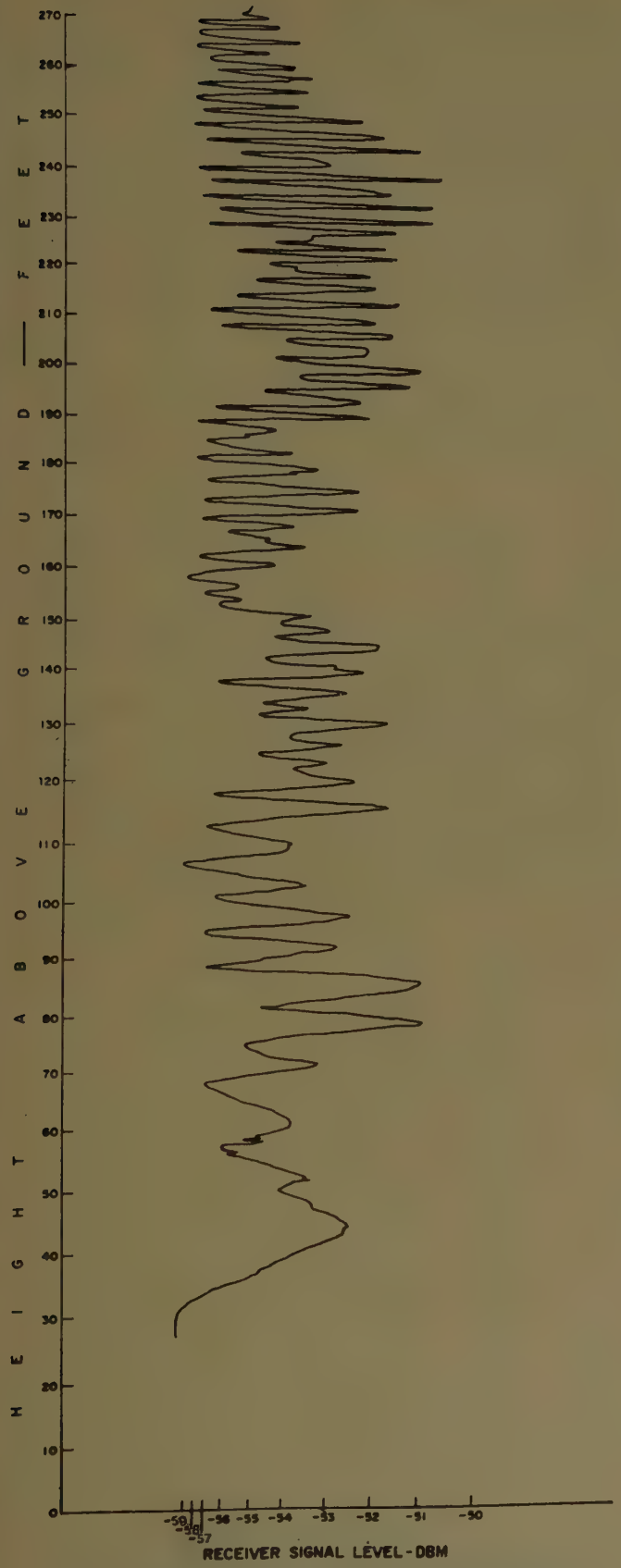


Fig. 12—Height-gain pattern Merrelltown site horizontal polarization. August 9, 1957. 11:00 A.M.

For the St. Edwards path, the spacing between the major lobes varied between 9 and 20 feet in an irregular fashion. An attempt was made to find the reflecting areas by choosing lobe spacings of 10, 15, and 20 feet at the appropriate receiver height. The location of the image antenna for each condition was found and the three dashed lines in Fig. 8 were drawn. Since these lines crossed the profile in a region which was not illuminated by the transmitter, it was felt that the complex path and the resulting signal strength curve did not permit ready identification of the reflecting areas.

For the Terrell Hill path, a broad smoothing of the height-gain curve indicates the presence of several lobes for the lower receiver heights with spacings of approximately 60 feet between nulls and three nulls with spacings of about 20 feet near the top of the tower. When the image transmitting antennas are located and the lines drawn to the appropriate place on the receiving tower, the dashed lines of Fig. 10 are obtained. These lines pass through the areas marked *C* and *D* which would seem to account fairly well for some of the major features of the signal strength curve. The many smaller variations are unexplained, as indicated earlier, and are thought to be due to scatter from the buildings along the path.

Examples of More Readily Identifiable Reflection Areas

Three paths were used on which the reflecting regions could be identified with some assurance. On the Merrelltown path there seemed to be a single reflection region, on the Barton Skyways path there seemed to be two principal reflecting areas, while on the East Base path there were several reflection areas.

On the Merrelltown path, the lobe structure as seen in Fig. 12 was most pronounced, ranging in magnitude from 3 to 6 db with equivalent reflection coefficients from 0.17 to 0.33. The lobes were not uniformly spaced but were closer together at greater receiver heights. When the locations of the image transmitter were found for several receiver elevations, the rays from the various image transmitter positions to the appropriate receiver heights all passed through or near a rounded portion of the profile as seen in Fig. 13. Visual observation showed that a road visible from both ends passed over this knoll

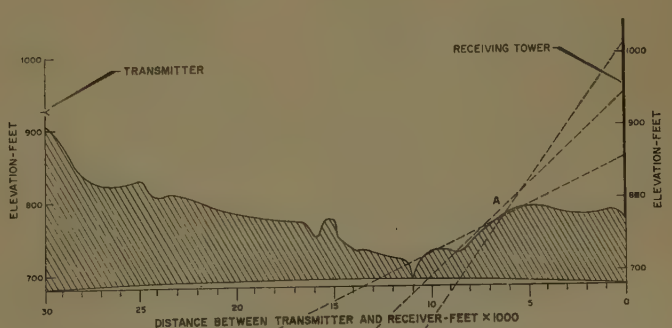


Fig. 13—Elevation profile Merrelltown to Balcones Research Center.

directly along the path. This indicated that this same general area served as a reflector for all receiver heights.

The Merrelltown path was therefore best suited for the time studies described later since the number of variables in the transmission path was reduced.

The Barton Skyway profile is shown in Fig. 14 and the associated height-gain curve is given in Fig. 15. A close inspection of this profile indicates two superimposed lobe structures with spacing of about 75 and 25 feet. The image transmitting antennas were located and lines were drawn from them to the middle of the tower. These two lines pass through two areas which would seem to meet the requirements for areas of reflection. The equivalent reflection coefficient was of the order of 0.2.

The East Base site, whose profile is shown in Fig. 16, had a height-gain curve with fairly well defined but irregularly spaced lobes as shown in Fig. 17. By identifying image transmitter locations with various possible spacings, several potential reflecting areas were noted. As found for the Merrelltown path, two knolls marked *C* and *D* appeared as potential reflectors while the flat area at *A* was a third possible source. The large lobes associated with reflection from *A* had a range of about 4 db while the smaller lobes associated with *C* and *D* had a range of 2 to 3 db. This corresponds to equivalent reflection coefficients of 0.23 and 0.15, respectively.

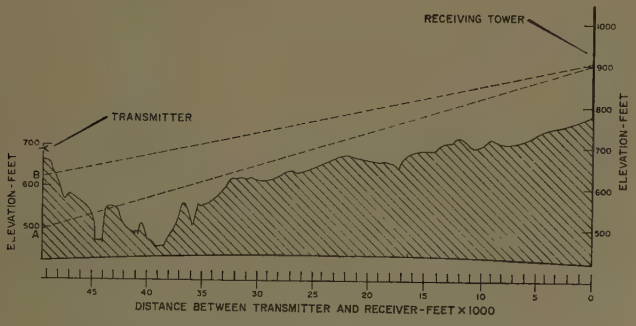


Fig. 14—Elevation profile Barton Skyway to Balcones Research Center.

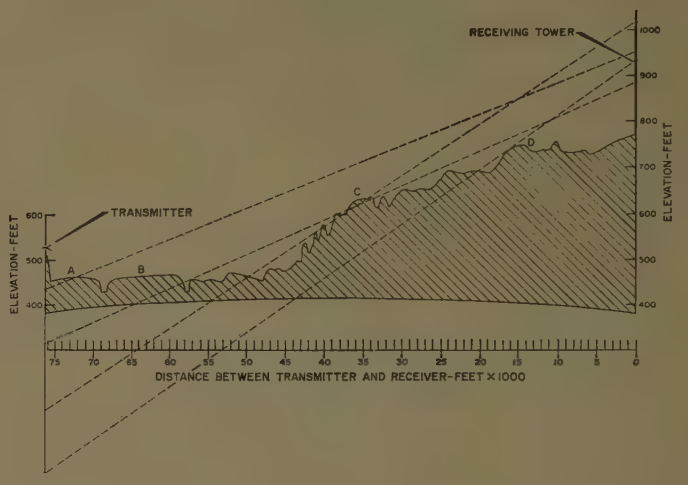


Fig. 16—Elevation profile East Base to Balcones Research Center.

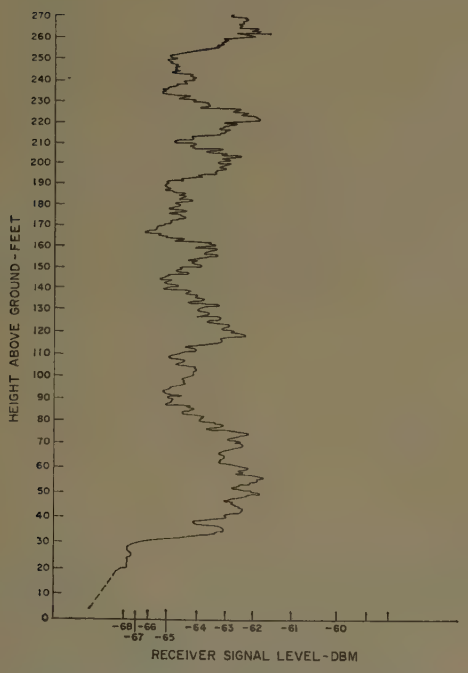


Fig. 15—Height-gain pattern Barton Skyway site horizonta polarization. July 26, 1957, 10:15 A.M.

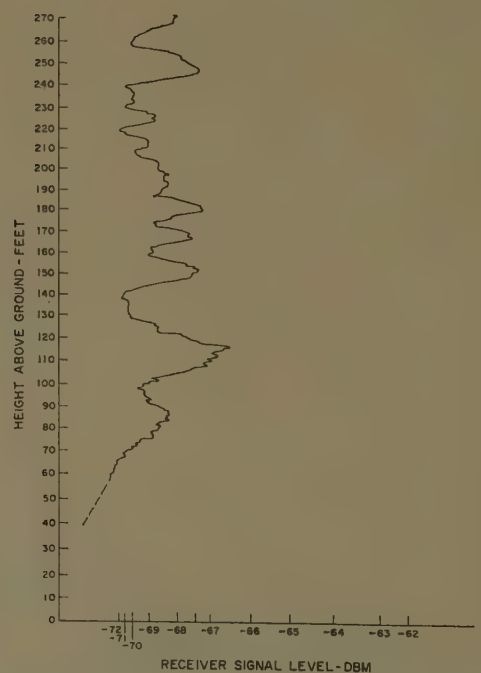


Fig. 17—Height-gain pattern East Base site horizontal polarization. August 8, 1957. 11:17 A.M.

Example of Curves Approximating Straight-Edge Diffraction

The profile for the Mt. Larson path is shown in Fig. 18 and the associated height-gain curve is shown in Fig. 19. The ridge shown at *A* is tree-covered and has an effective height of about 30 feet above the ground. A line through the top of the trees passes 30 feet above ground at the receiving tower. Using these dimensions and the diffraction theory, a height-gain curve was calculated and is shown in Fig. 20 with the measured height-gain curve smoothed to remove the smaller variations. The agreement between the measured and calculated curves was thought to be very good.

The Four Corners profile and the associated height-gain curve are shown in Figs. 21 and 22, respectively. The ridges were covered with trees which increased

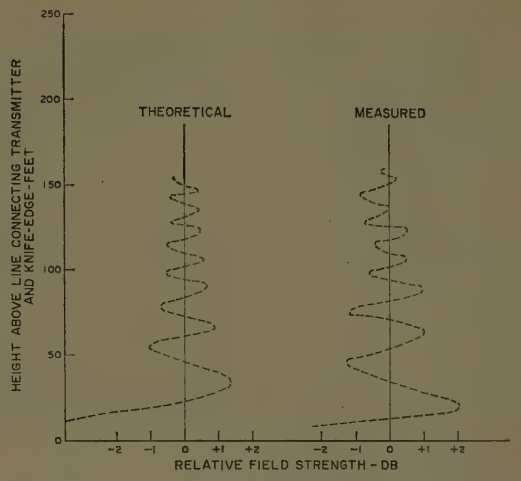


Fig. 20—Relative field strength at receiver vs height above line connecting transmitter and knife edge.

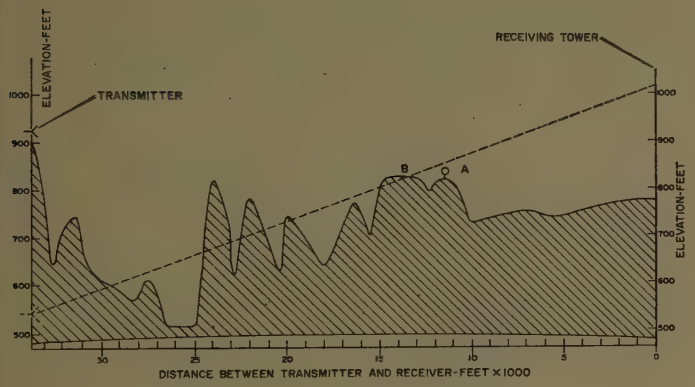


Fig. 18—Elevation profile Mt. Larson to Balcones Research Center

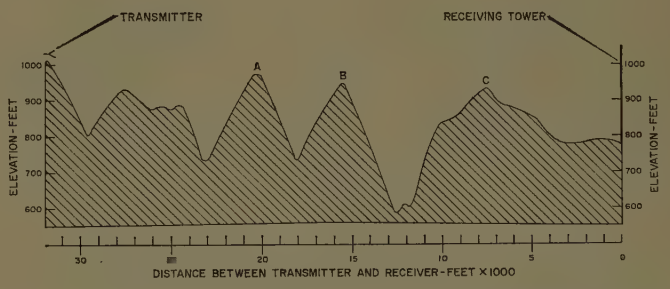


Fig. 21—Elevation profile Four Corners to Balcones Research Center.

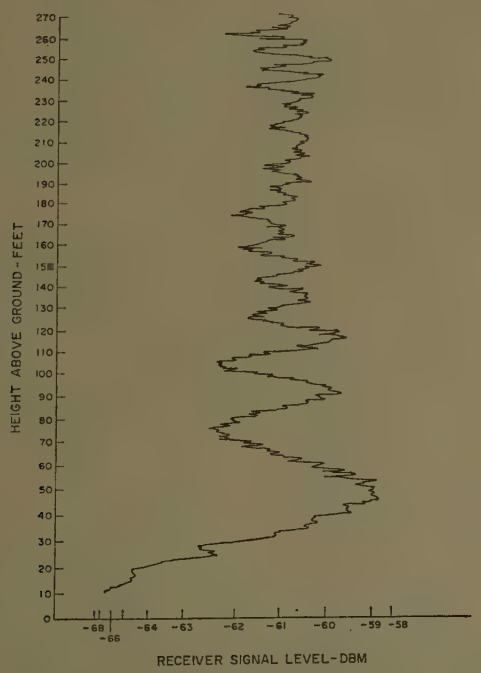


Fig. 19—Height-gain pattern Mt. Larson site horizontal polarization. July 30, 1957. 11:10 A.M.

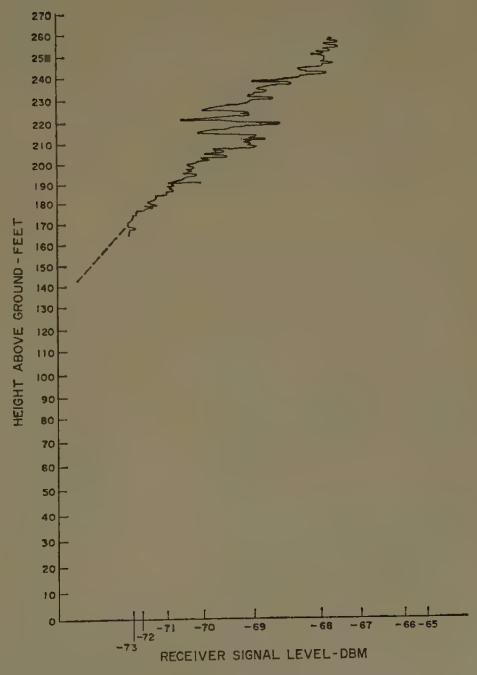


Fig. 22—Height-gain pattern Four Corners site horizontal polarization. August 2, 1957. 11:45 A.M.

their effective height. Only the upper portion of the tower is in the line of sight and the height-gain curve is thought to be that associated with the first lobe of the diffraction pattern with insufficient height reached to attain the first maximum.

Variation of the Height of Nulls

The Merrelltown transmission path was chosen for time studies of the variation of the height of the nulls as a function of time since the reflection area was most clearly defined.

For the time studies, height-gain patterns were taken at intervals of 20 to 30 minutes and vertical wet- and dry-bulb temperature measurements were taken hourly. Six time studies were made, two around the sunset interval, two around sunrise, and two over 24-hour periods. In all cases there were sharp shifts in the heights of the nulls around sunrise and sunset.

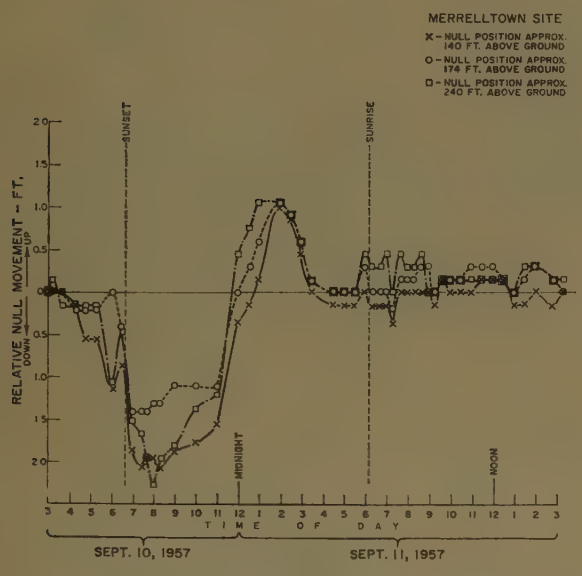


Fig. 23—Null movement vs time of day.

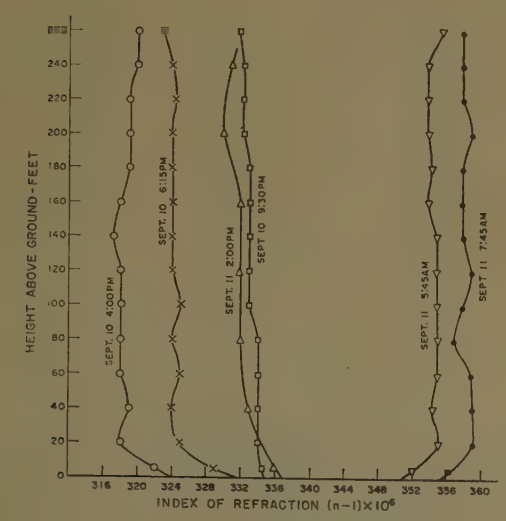


Fig. 24—Index of refraction vs altitude. Balcones Research Center, September 10–11, 1957.

In a run on September 10 and 11, 1957, the positions of nulls in the vicinity of 140, 174, and 240 feet on the tower were followed, and this movement is shown in Fig. 23. Several refractive index height profiles taken during the measurement period are shown in Fig. 24. It is known that the refractive index varies rapidly with time and spacing and the values determined from the wet- and dry-bulb thermometers provide a great deal of smoothing. Greater precision at a single point would not, however, be of much help since the horizontal variation of index of refraction would also be large.

The second 24-hour run was made on October 17 and 18, 1957, and is shown in Fig. 25 with selected refractive index soundings during the period shown in Fig. 26.

As indicated earlier, the change in the position of the nulls may be predicted if the average index of refraction along the direct and reflected paths is known. For this purpose the average index over the height intervals through which the direct and reflected wave travels were calculated for each meteorological sounding. A comparison was made of successive soundings to give a

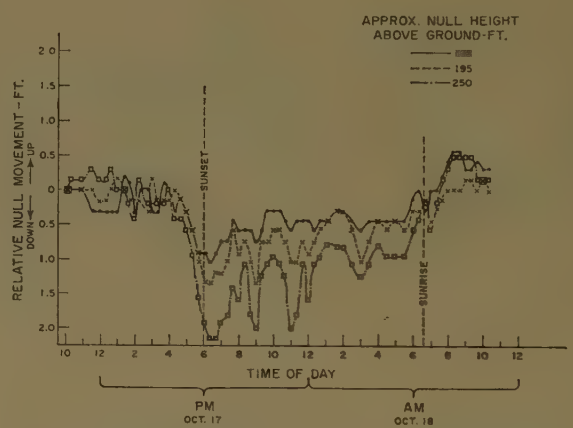


Fig. 25—Null movement vs time of day. Merrelltown site vertical polarization. October 17–18, 1957.

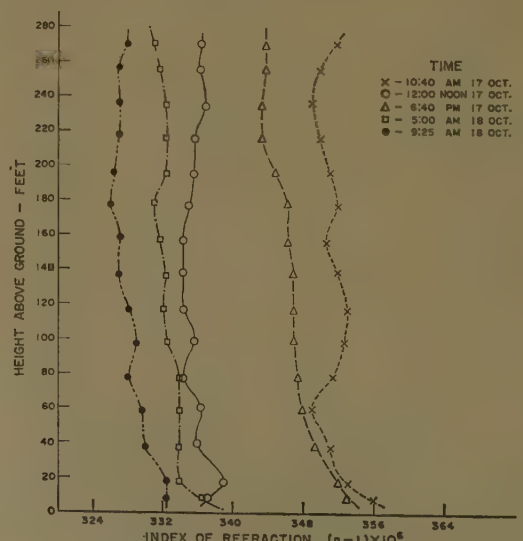


Fig. 26—Index of refraction (n) vs altitude. Balcones Research Center, October 17–18, 1957.

theoretical change in the nulls. These theoretical changes are plotted in Fig. 27 as a function of the measured changes. For an exact correlation, the points would fall on a 45° line. There is considerable scatter in the points in Fig. 27 but this is to be expected because of the assumptions which were made. The procedure for calculating the average index over the paths is poor since it requires horizontal stratification of meteorological parameters even over the rolling terrain. In spite of this broad assumption, the direction of change was in nearly every instance in agreement with the direction predicted from the index-of-refraction profile variations.

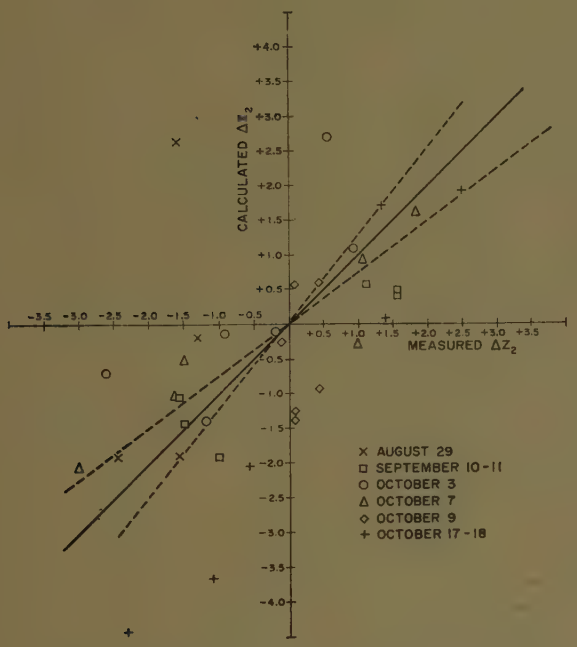


Fig. 27—Measured vs calculated null movements.

Conclusion

The study of height-gain functions, although made over rather complex terrain, showed that reflection areas could be identified which seemed to fit the requirements for sources of interfering signal. The observed signal-height curves could be explained in part by these reflection areas, with the number of areas being different for different profiles. Some fine scale variations were thought to be due to random reflectors such as buildings.

The equivalent reflection coefficient ranged from 0.15 to 0.33. Measurements over some wooded hills agreed with the equivalent reflection coefficient of 0.17 measured by Sharpless [3] over somewhat similar terrain.

Time studies of the height of nulls showed diurnal variations of 3 to 4 feet over the paths used. Analysis indicated some correlation between the measured null movement and that calculated from index of refraction soundings.

The analysis of the height-gain function is a complex one and it is not anticipated that a one to one agreement with prediction from profile and meteorological measurement will ever be attained. It is felt, however, that data and analyses such as those presented here should be used in predicting the general propagation characteristics of low-angle radio waves.

Bibliography

[1] H. T. Tomlinson, "Effects of Topography and Refractive Index on Low-Angle 3 CM Radio Waves," Ph.D. dissertation, Univ. of Texas, Austin; 1958. Available from University Microfilms, Ann Arbor, Mich.

[2]D. E. Kerr, ed., "Propagation of Short Radio Waves," M.I.T. Rad. Lab. Ser., McGraw-Hill Book Co., Inc., New York, N. Y., vol. 13; 1951.

[3] W. M. Sharpless, "Ground Reflection Coefficient Experiments at *X*-Band," Bell Telephone Labs., Red Bank, N. J., Rept. No. MM-44-160-260; December, 1944.

Electron Densities of the Ionosphere Utilizing High-Altitude Rockets*

O. C. HAYCOCK†, J. I. SWIGART†, AND D. J. BAKER‡

Summary—The problem of determining the electron densities in the *E*-region of the ionosphere is approached by using 6-mc pulse transmissions from a rocket to several ground receiving stations.

A logical and complete development, using dyadic techniques, is given for obtaining the propagation constant of the dissipative, anisotropic ionosphere. Special cases of the magneto-ionic formulas are given, and comparison of the ionosphere with a distributed-constant transmission line is made.

In a nondissipative ionosphere, formulas are developed establishing the relationship between the effective electron density and the relative transmission delay of the 6-mc pulse.

A description of the University of Utah's vertical incidence experiment is given in which a 6-mc pulse from an airborne transmitter is received simultaneously at several ground receiving stations.

The relative 6-mc time-delay data from three Aerobee high-altitude rockets launched from Holloman Air Development Center on July 1, 1953, November 3, 1953, and June 13, 1956, were obtained and, from these, electron density was calculated. Curves showing the profile of electron density as a function of altitude as calculated both during the rocket ascent and descent are presented. The curves indicate a general increase of electron density throughout the *E*-region, rising from nearly zero at 85 km to a maximum of about 2×10^{11} electrons/m^3. The maximum altitude attained by the rockets allowed exploration up to 137 km above sea level.

THE OBJECT of the research reported by this paper is the determination of the electron density in the *E*-region of the earth's ionosphere, and its variations in both time and space.

The first part of the report is a new derivation of the equations of ionospheric transmission. Most derivations imply a negative capacitance in the ionosphere, whereas free electrons would ordinarily be considered inductive. The criticism is conceptual rather than with the mathematical results. Baker then prepared the new derivation, following more nearly the electromagnetic-wave and transmission-line theory common in electrical engineering, which we feel reduces the conceptual difficulties. The second part shows the results of the application of the transmission equations to data obtained over a period of years to obtain electron-density.

The method of attack toward attainment of the objective is the determination of electron densities by means of radio-frequency pulses transmitted through the *E*-region from an Aerobee high-altitude sounding rocket to the ground (Fig. 1).[1] A pulse transmitted through the region at a frequency somewhat greater than the critical frequency of the *E*-layer suffers anomalous dispersion in that its phase velocity exceeds that of light in free space, while its signal velocity is less than that in free space. A frequency of 6 mc has been chosen for the perturbing pulse. The transmission time of the perturbing pulse is compared with that of a synchronized UHF pulse, also transmitted from the rocket. A frequency near 600 mc has been chosen for the reference pulse because such pulses are negligibly retarded by passage through the ionosphere. Large scale space variations could be obtained by receiving the downward transmissions at a number of sites separated by relatively large distances. The signal velocity as a function of altitude and the ray path must be established from the relative retardation times recorded throughout the rocket flight. The effective electron density can then be calculated by means of magneto-ionic theory thus giving the distribution of electron density in the *E*-region as required.

In order to determine the manner of propagation of a plane wave through the ionosphere, it is necessary to consider the motions of the charges in the medium. Because of the great mass of a positive ion compared with that of an electron, the positive ions do not obtain sufficient velocities to contribute significantly to the current density produced by the passing waves. Hence, only electron flow will be considered.

A plane electromagnetic wave is considered to be propagated in the positive Z-direction of a right-handed Cartesian coordinate system through an ionized medium containing N electrons per unit volume. The electron density is assumed to be homogeneous over distances comparable with a wavelength (Fig. 2).

The earth imposes a static magnetic field $\boldsymbol{B}_e$ which may be resolved into components B_l parallel to and B_t transverse to the Z-axis.

The classical equation for the force on the charge per unit volume may be written

$$-\frac{m}{e}\frac{d\boldsymbol{J}}{dt} = -Ne\boldsymbol{E} + \boldsymbol{J}\times\boldsymbol{B}_e + \frac{m\nu}{e}\boldsymbol{J}, \qquad (1)$$

where $\boldsymbol{E}$ is the electric field intensity vector of the traveling wave, ν is the number on collisions per second per

* Manuscript received by the PGAP, December 9, 1957; revised manuscript received, March 4, 1959. This research has been carried out both at the University of Utah, Salt Lake City, Utah, and at Holloman ADC under Contract No. AF 19(604)-384 from the Geophys. Res. Directorate, Air Force Cambridge Res. Cen. Dr. Leon B. Linford served as project director until his death on March 12, 1957.

† University of Utah, Salt Lake City, Utah.

‡ USAF, Geophys. Res. Directorate, AFCRC, Cambridge, Mass.

[1] L. B. Linford, "Determination of the Electron Density in the Ionosphere," University of Utah, Salt Lake City, Progress Rept. No. 1, Contract No. AF 19(604)-384, pp. 1–30; September, 1952.

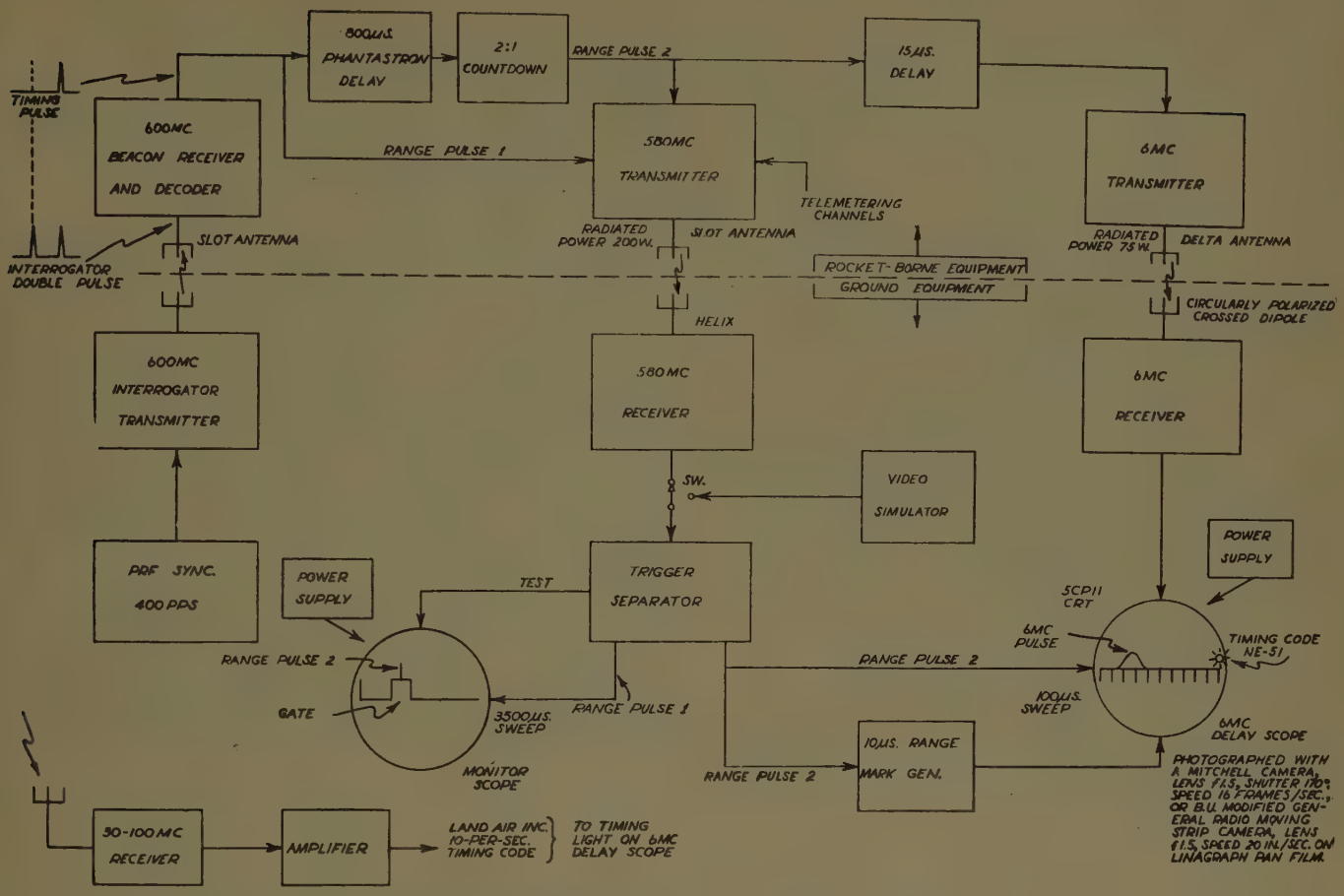


Fig. 1—Block diagram of 6-mc vertical incidence experiment.

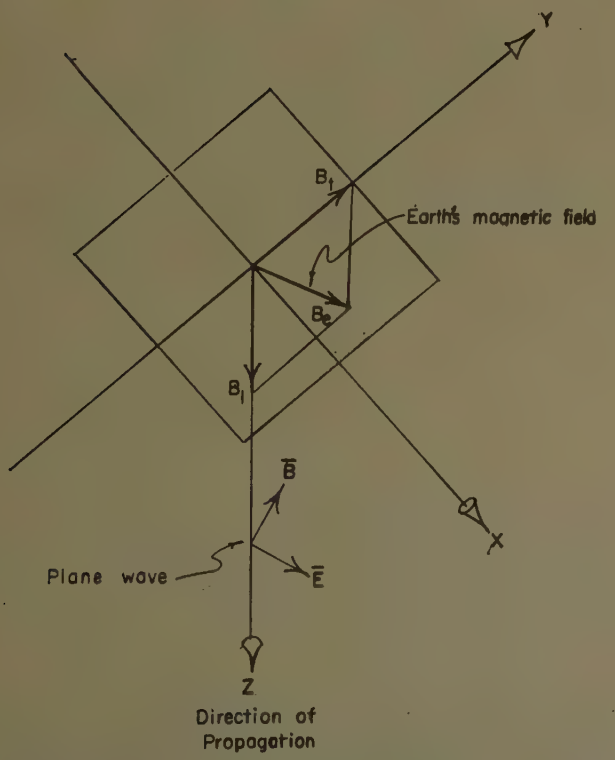


Fig. 2—The propagation of a plane wave through the ionosphere.

particle of the electron with other particles of the medium, e and m are the charge and mass, respectively, of the electron, and J is the total current density.

The force caused by the magnetic field of the perturbing wave is neglected, the Lorentz polarization term is omitted,[2] and it is assumed that any higher order effects caused by nonlinearity of the medium and "electron bunching" are negligible.[3]

$$NeE = \frac{m}{e}(j\omega + \nu)J + J \times B_e, \tag{2}$$

where ω is the angular frequency of the probing wave. Eq. (2) may be expressed in tensor form, and when the current density vector is expressed as a linear function of the electric field intensity vector, there results

$$J = \Sigma \cdot E, \tag{3}$$

where Σ, a tensor of second rank, is the nonian form of the conductivity dyadic of the medium represented by the matrix,

[2] C. Darwin, "The refractive index of an ionized medium," *Proc. Roy. Soc.*, vol. 182A, pp. 152–166; December, 1943.

[3] C. O. Hines, "Higher-order approximations in ionospheric wave propagation," *J. Geophys. Res.*, vol. 58, pp. 95–98; March, 1953.

$$\Sigma = \frac{Ne}{C(C^2 + B_e^2)} \begin{bmatrix} C^2 & -B_1C & B_tC \\ B_1C & C^2 + B_t^2 & B_tB_1 \\ -B_tC & B_tB_1 & C^2 + B_1^2 \end{bmatrix}, \quad (4)$$

where

$$C = \frac{m}{e}(j\omega + \nu). \quad (5)$$

The induced polarization of the ionosphere caused by the oscillating electrons is

$$P = \frac{1}{j\omega} \Sigma \cdot E = \epsilon_0 K \cdot E, \quad (6)$$

where ϵ_0 is the permittivity of free space and K is the electric susceptibility dyadic of the ionosphere. Only when collisions are neglected ($\nu \ll \omega$) does this dyadic become Hermitian. The Hermitian or non-Hermitian character of the susceptibility dyadic may be taken as a criterion for indicating whether the medium is non-dissipative or dissipative, respectively.[4]

The electric displacement is given by

$$D = \epsilon_0 E + P = \epsilon_0 \varepsilon \cdot E, \quad (7)$$

where

$$\varepsilon = I + \frac{1}{j\omega\epsilon_0} \Sigma \quad (8)$$

is the permittivity dyadic of the medium, and I is the unit dyadic.

The substitution of (7) into Maxwell's equations for the time periodic, macroscopic field vectors, and the subsequent elimination of the magnetic field intensity vector yields the wave equation,

$$\nabla^2 E - \nabla(\nabla \cdot E) = -\frac{\omega^2}{c^2} \varepsilon \cdot E, \quad (9)$$

where c is the velocity of light in free space. This equation has a plane traveling-wave solution of the form $\exp j\omega t - \gamma z$ only for the eigenvalues of the propagation constant given by equating the determinant of the characteristic matrix,

$$-\frac{\omega^2}{c^2} E - \begin{bmatrix} \gamma^2 & 0 & 0 \\ 0 & \gamma^2 & 0, \\ 0 & 0 & 0 \end{bmatrix} \quad (10)$$

to zero. The square of the propagation constant will be real only when this matrix is Hermitian (nondissipative case).

The characteristic equation gives a quadratic in γ^2; the quadratic formula can, therefore, be invoked to yield for the square of the propagation constant

$$\gamma^2 = -\frac{\omega^2}{c^2} + \frac{N^2e^2\omega\mu_0}{j\omega\epsilon_0 C(C^2 + B_e^2) + Ne(C^2 + B_1^2)} \{jNeC - \omega\epsilon_0 (C^2 + \tfrac{1}{2}B_t^2) \pm \sqrt{(j\omega\epsilon_0)^2(C^2B_1^2 - B_t^2/4) + j2Ne\omega\epsilon_0 C + N^2e^2}\}, \quad (11)$$

where μ_0 is the permeability of free space.

The square of the ordinary ray propagation constant for the case of no magnetic field is

$$\gamma^2 = -\frac{\omega^2}{c^2} + \frac{Ne^2\omega\mu_0}{m(\omega - j\nu)}. \quad (12)$$

The ionosphere can be treated as an equivalent high-pass filter. This is shown in Appendix A.

The time delay of a 6-mc probing pulse relative to a synchronized UHF pulse is the result of both a reduced signal velocity and propagation over a ray path which differs from line of sight.

Assuming that there is no horizontal inhomogeneity and neglecting the terrestrial magnetic field, the ionosphere may be divided into horizontal layers, each of thickness Δh_i, so that each layer is homogeneous within itself. The transmission time for a signal to traverse a path from the rocket in the qth layer to the ground is

$$t_q = \sum_{i=0}^{q} \frac{\Delta s_i}{u_{gi}}, \quad (13)$$

where Δs_i is the length of the straight path in the ith layer, and u_{gi} is the signal velocity in the ith layer. By applying Snell's law to the geometry of ray propagation through the ith layer, (13) becomes

$$t_q = \sum_{i=0}^{q} \frac{\Delta h_i n_i}{u_{gi}\sqrt{n_i^2 - x_q^2/(ct_q)^2}}, \quad (14)$$

where n_i is the index of refraction in the ith layer, and x_q is the ground projection distance from the rocket transmitter to the ground receiver.

By separating the last term in the series, (14) may be solved for n_q^2. Using $\beta = n\omega/c$, the resulting phase constant may be substituted ($\nu \ll \omega$) into (12) to yield

$$N_q = \frac{m\omega^2\epsilon_0}{e^2} \left\{ 1 - \frac{x_q^2}{(ct_q)^2} - \frac{\Delta h_q^2}{\left[ct_q - \sum_{i=0}^{q-1} \frac{\Delta h_i}{\sqrt{n_i^2 - x_q^2/(ct_q)^2}} \right]^2} \right\}, \quad (15)$$

when solved for the electron density of the qth layer.

An alternative formula may be found from (14) by finding the time Δt_q taken for the signal to traverse the qth layer.[5] When solved for the phase constant, which is substituted into (12), the resulting equation gives

[4] P. C. Clemmow and J. Heading, "Coupled forms of the differential equations governing radio propagation in the ionosphere," *Proc. Camb. Phil. Soc.*, vol. 50, p. 319; 1954.

[5] J. R. Lien, R. J. Marcou, J. C. Ulwick and D. R. McMorrow, "Ionosphere research with rocket-borne instruments," in "Rocket Exploration of the Upper Atmosphere," Interscience Publishers, Inc., New York, N. Y., p. 224; 1954.

$$N_q = \frac{m\omega^2\epsilon_0}{e^2}\left\{1 - x_q^2/(ct_q)^2 - (\Delta h_q/c\Delta t_q)^2\right\} \qquad (16)$$

for the electron density of the qth layer. The time t_q is related to the recorded relative delay time t_d by

$$t_q = t_d + \frac{1}{c}\sqrt{h_q^2 + x_q^2}. \qquad (17)$$

The first rocket from which the University of Utah, in conjunction with the Air Force Cambridge Research Center, received usable 6-mc relative delay data was USAF Aerobee No. 39, launched at 1052 hours, MST, from Holloman Air Development Center, N. Mex. on July 1, 1953. The effective electron density profile calculated from these delay data using either (15) or (16) is given for both the rocket ascent and descent in Fig. 3.

A second rocket from which downward 6-mc transmission data were obtained was USAF Aerobee No. 45, launched at 1115 hours, MST, November 3, 1953. The results from this flight are given in Fig. 4.

Fig. 5 presents the results calculated from a third rocket, USAF Interim Aerobee Hi No. 67, launched at Holloman Air Development Center at 1352 hours, MST, June 13, 1956.

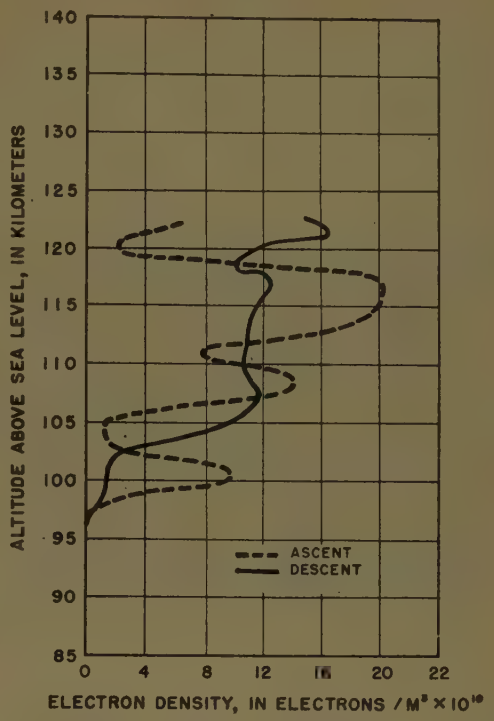


Fig. 4—Effective electron density calculated from the Army-2 delay data from USAF Aerobee no. 45, launched at Holloman Air Development Center, N. Mex., 1115 hr. MST, November 3, 1953.

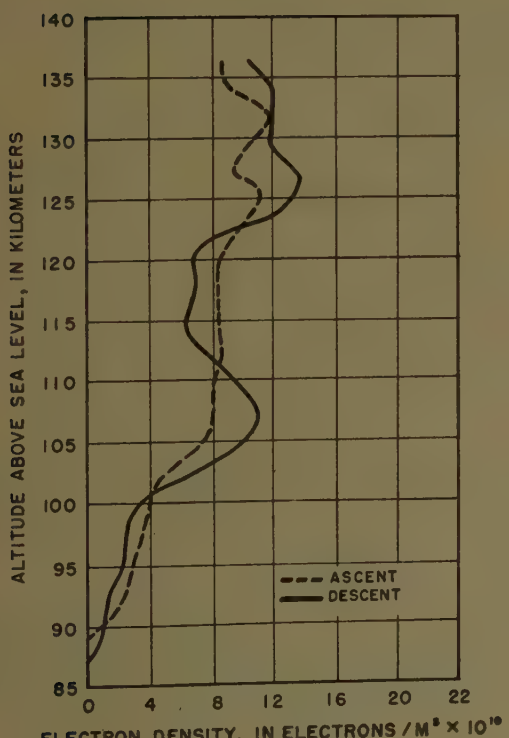


Fig. 3—Effective electron density calculated from the skillet knob delay data from USAF Aerobee no. 39, launched at Holloman Air Development Center, N. Mex., 1052 hr. MST, July 1, 1953.

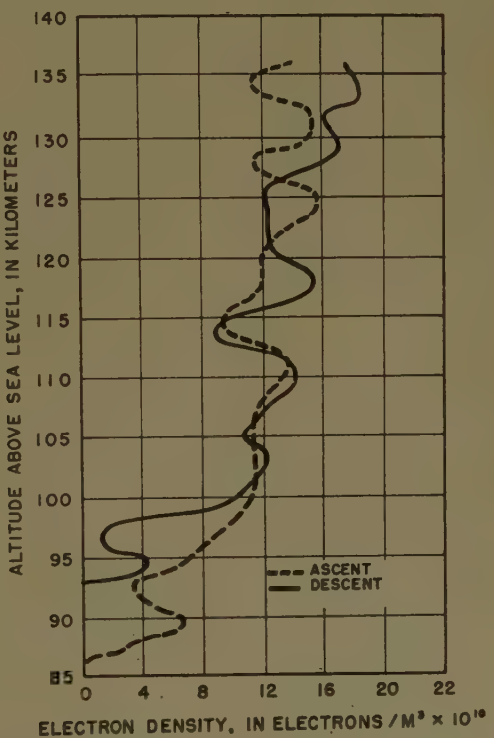


Fig. 5—Effective electron density calculated from the skillet knob delay data from USAF Aerobee no. 67, launched at Holloman Air Development Center, N. Mex., 1352 hr., MST, June 13, 1956.

Conclusion

The following conclusions have been reached from the curves of effective electron density, calculated from the relative transmission delay of a 6-mc pulse in its downward passage through the E-region of the ionosphere:

1) The electron density has a definite general increase with altitude throughout the E-region, which has here been explored to a height of 137 km above sea level.

2) There exist striations, which are sometimes very marked, throughout the E-region.

3) A pronounced difference occurs in the profile of electron density during the rocket descent as compared with that during the rocket ascent. This could be due to the changed transmission path, indicating a horizontal inhomogeneity (patchiness), and to rapid time variations.

4) There is a very pronounced difference in the vertical structure of the ionosphere from one rocket firing to the next. This could be due to variable striations superimposed upon a somewhat stable Chapman layer.

5) The values of electron density range from nearly zero below 85 km to an observed maximum of about 2×10^{11} electrons/m^3.

A comparison with the results from other experiments will now be given. Lien, *et al.*[6] have computed an electron density profile curve from relative time delay data of a 4.87-mc pulse transmission from a ground transmitter to a receiver in an Aerobee rocket launched at Holloman ADC on June 26, 1953. These authors indicate a much more pronounced bifurcation of the E-region; however, the maximum electron density of nearly 2×10^{11} electrons/m^3 is the same.

Seddon and Jackson[7] have determined the electron density profile using Viking rockets launched at White Sands Proving Ground, N. Mex. The experiment consists in phase velocity measurements accomplished by transmitting a continuous-wave signal, selected to obtain the data, and a sixth harmonic, used as a reference, from the rocket to receiving stations on the ground. Beat notes are obtained from which the indexes of refraction and, subsequently, the electron densities can be calculated. While the results indicate somewhat of an over-all similarity, with the absence of a marked bifurcation, they do not show striations. The maximum electron densities reached are about 2×10^{11} electrons/m^3.

Berning[8] has calculated the electron densities from rocket flights using radio-Doppler techniques. This method, using the DOVAP radio-Doppler electronic system of tracking, is to compare the radial velocities determined by vacuum trajectory formulas with those actually measured. The differences are converted to ionization densities at the levels through which the rocket is passing. The electron density profile as calculated from a Viking launched at White Sands Proving Grounds on December 15, 1952 indicates a general bifurcation of the E-region, but no fine structure is apparent. A maximum electron density of about 1.5×10^{11} electrons/m^3 is reached in the E-region.

Acknowledgment

The authors desire to express their appreciation for the excellent work of Dr. Leon B. Linford, who was director of the project until the time of his death, and who contributed greatly to its success.

They also wish to acknowledge their appreciation to the Geophysics Research Directorate, AFCRC, and to those who have coordinated their effort at AFCRC, Dr. W. Pfister, J. C. Ulwick, Dr. M. D. O'Day, and J. R. Lien, as well as to the staff members of the University of Utah who have worked on this contract at various times: Professors E. C. Madsen, C. L. Alley, and C. D. Westlund, and finally to all those undergraduate and graduate students who helped develop the apparatus to the point at which the reported data could be obtained, and who aided in taking it.

Appendix A

An equivalent transmission line with infinitesimal distributed constants and a propagation constant given by (12) is shown in Fig. 6.

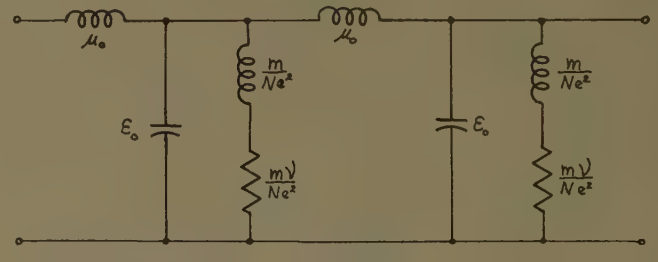


Fig. 6—Equivalent transmission line of ionosphere.

If collisions are neglected, the characteristic impedance of this network is

$$Z_0 = \sqrt{\frac{\omega^2 m\mu_0}{\omega^2 m\epsilon_0 - Ne^2}}. \tag{18}$$

The ionosphere in this case behaves like a high-pass filter, with Z_0 becoming infinite at the critical frequency,

$$f_c = \frac{1}{2\pi}\sqrt{\frac{Ne^2}{m\epsilon_0}}. \tag{19}$$

For frequencies below f_c, Z_0 is inductive and the propagation constant is real; above the critical frequency, Z_0 is real and γ is imaginary.

[6] Lien, *et al.*, *op. cit.*, p. 232.

[7] J. C. Seddon and J. E. Jackson, "Absence of bifurcation in the E-layer," *Phys. Rev.*, vol. 97, pp. 1182–1183; 1955; and J. C. Seddon, A. D. Picker and J. E. Jackson, "Continuous electron density measurements up to 200 km," *J. Geophys. Res.*, vol. 59, pp. 513–524; 1954.

[8] W. W. Berning, "Charge densities in the ionosphere from radio doppler data," in "Rocket Exploration of the Upper Atmosphere," Interscience Publishers, Inc., New York, N. Y., p. 272; 1954.

A Scatter Propagation Experiment Using an Array of Six Paraboloids*

LORNE H. DOHERTY†

Summary—**Using an antenna system whose aperture could be varied in four-foot steps between 4 and 24 feet, aperture-to-medium coupling loss measurements have been made on a 2720-msec, 216-mile path. These measurements reveal an intrinsic variability in the scattering mechanism which is not accounted for in most current theories. Diversity and fading-rate measurements were also made. A simple mathematical model of the diffracted field yields calculated values of the normal component of the wind which agree well with the measured wind. Calculated and measured values of fading rate are also seen to be in good agreement. An estimate is made of the turbulent wind velocity.**

Introduction

A 2720-msec radio link was operated between Ottawa and Toronto, a distance of 216 miles, for a period of 20 months (from April, 1955 to December, 1956). The observations of signal level and details of the equipment were reported in a previous paper.[1] During part of this 20-month period, and for four months following termination of the routine observations, experiments were performed with a special receiving antenna system. This antenna system consisted of six four-foot-diameter paraboloids mounted side-by-side horizontally, at right angles to the propagation path. Using the antennas individually, measurements were made of the correlation of signal amplitudes across the path. Any number of these antennas could be paralleled to form, with proper phasing, an antenna array whose aperture could be varied in four-foot steps between four feet and 24 feet.

Experiments with this antenna system have yielded results on aperture-to-medium coupling loss, on fading rates as a function of antenna size, on correlation distances normal to the path, on the effect of the average wind on fading rates and antenna gain, and on turbulent wind velocity.

The Antenna Array

A photograph of the six paraboloids is shown in Fig. 1. The centers of the antennas were about six feet from ground level and were mounted with about two inches between their rims. The antenna faces were laid out along a line at right angles to the propagation path. A waveguide double dipole feed was used, but a transition to coaxial line was placed immediately behind the paraboloids. Flexible 50-ohm lines (RG-8/U) were used to feed the signals from the individual antennas to a housing behind the center of the array. These lines were cut to equal length. A coaxial line stretcher was placed in each of these lines and mounted in the central housing. A matching and paralleling unit was mounted in the housing. This unit had six input connectors (one for each antenna) and an output connection to a receiver. If fewer than six antennas were to be paralleled, a shorted stub was placed on the unused connectors. A variable stub, calibrated in terms of the number of antennas paralleled, completed the unit. A voltage standing-wave ratio of better than 1.3 was obtained with any antenna combination. A photograph of the housing is shown in Fig. 2.

Fig. 1—Photograph of the array of six paraboloids.

The array was placed just a few feet back from the edge of a cliff. Consequently, it was impossible to make the necessary phase adjustments with a nearby antenna providing a signal. The nearest accessible point for this calibrating antenna was some 15 miles away over water. The signal received from this distance was almost invariably fading over a 10- to 20-db range. On a few occasions (usually during rain or below-freezing temperatures when outdoor working conditions were at their worst), a comparatively steady signal was received and the phase adjustments of the antennas could be checked.

The initial phase adjustments were made using the scatter signal transmitted from Ottawa as a signal source. Because of the aperture-to-medium coupling loss, the observed signal level was not sensitive to the line-stretcher adjustment when more than three antennas were paralleled. Consequently the following procedure was adopted. Two antennas at one end of the array were paralleled. The line stretcher of one was adjusted to maximize the median signal received during a period which varied between 0.5 minute and 3.0 minutes. The signal from another antenna was observed during this same period to determine what changes in signal level were independent of the line stretcher ad-

* Manuscript received by the PGAP, March 25, 1959.
† Radio and Elect. Engrg. Div., Natl. Res. Council, Ottawa, Can.
[1] L. H. Doherty and G. Neal, "A 215-mile 2720-msec radio link," to be published in IRE Trans. on Antennas and Propagation.

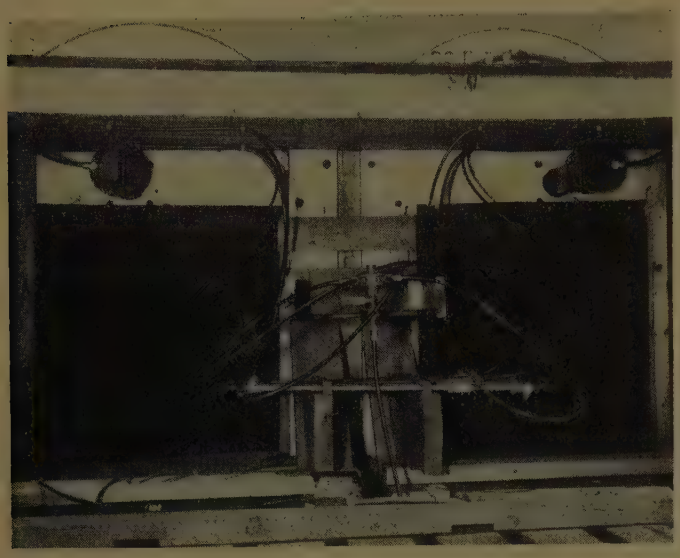


Fig. 2—Photograph of the interior of housing containing paralleling unit and line stretchers.

justments. Following this adjustment, the second and third antennas were paralleled and the line stretcher associated with the third antenna adjusted for maximum signal. This process was continued for all six antennas, each antenna being phased in conjunction with that immediately before it in the array.

A considerable degree of confidence in this method of phasing the antennas developed, and it was the method used exclusively in the latter half of the project. The most likely source of phasing error would be a cumulative one resulting in a progressive change in phase down the length of the array. One method of guarding against this possibility was by starting at the opposite end of the array and paralleling the paraboloids in pairs as before. No significant change in line-stretcher positions was ever observed when this experiment was performed. In addition, when any three antennas were paralleled, it was never possible to increase the signal level by a line-stretcher adjustment.

Some variation in the electrical length of the cables was observed and the phasing was checked periodically. This was not, however, a serious problem. Small variations in cable loss, frequently associated with the cable connectors, gave more trouble, and necessitated the replacement of at least one cable each month. The relative gains, including cable loss, of the individual antennas were checked each time the antenna system was used.

With the phasing accomplished as described, measurements of gain were made using a signal source placed along the path at a distance of 15 miles. This is a line-of-sight path, but as has already been mentioned, because of the almost continuous superrefraction over Lake Ontario,[2] the signal fluctuates over a 10 to 20-db range in a matter of one to five minutes. Table I lists the results of some gain measurements made during periods of relatively stable signal levels. The average gains should be compared with the plane-wave gains which would be obtained with perfect phasing of antennas of equal gain. The difference in measured gain between the lowest and highest figure is 1.6 db in the worst case, but the average agrees well with the calculated values, and there is no evidence of an improper phase adjustment.

Further evidence of the accuracy of the described method of phase adjustment may be found in the scatter signal gain measurements. In the results of the next and later sections, there are gain measurements yielding values in close agreement with the theoretical plane wave values.

TABLE I

PLANE-WAVE GAIN RELATIVE TO ONE ANTENNA (DB)

Number of Antennas Paralleled	Measured				Average	Calculated
	February 29	March 19	March 22	April 16		
2	3.5	2.7	4.0	3.0	3.3	3.0
3	4.3	5.1	5.6	4.0	4.7	4.8
4	5.6	6.7	6.8	6.3	6.4	6.0
5	6.5	7.4	7.3	7.2	7.1	7.0
6	—	7.7	7.8	7.7	7.7	7.8

PRELIMINARY MEASUREMENTS OF ANTENNA GAIN

Some preliminary measurements of relative antenna gain on the scatter signal from Ottawa were made during the latter half of June, 1955. In these experiments, no reference antenna was used. One, two, to six antennas were connected to the receiver in turn. The signal was recorded for a period of three minutes with a time constant of four seconds. The median signal level for each antenna arrangement was then estimated from the recording. The results from 18 such experiments are summarized in Fig. 3. These experiments were performed at a different site than the rest of the experiments covered in this paper.[1]

In Fig. 3, the gain is plotted relative to the gain of one antenna. Each curve is an average of two, three, or four observations as indicated in brackets after the data. One experiment involving a single three-minute observation for each antenna arrangement took about 25 minutes to perform.

The large variation in realized gain is immediately apparent from these curves. There would also appear to be a diurnal variation, higher gains being realized in the morning than in the afternoon. The afternoon observations show the lowest gain figures of any made with this array, and it is possible that the site had an adverse effect on the observations.

The results of a further set of preliminary measurements are shown in Fig. 4. In this case, one antenna of the array was used as a reference antenna, and conse-

[2] A. D. Hood and L. H. Doherty, "Radar Propagation on Lake Ontario," Natl. Res. Council of Canada, Ottawa, REED Rept. ERA-321: March, 1957.

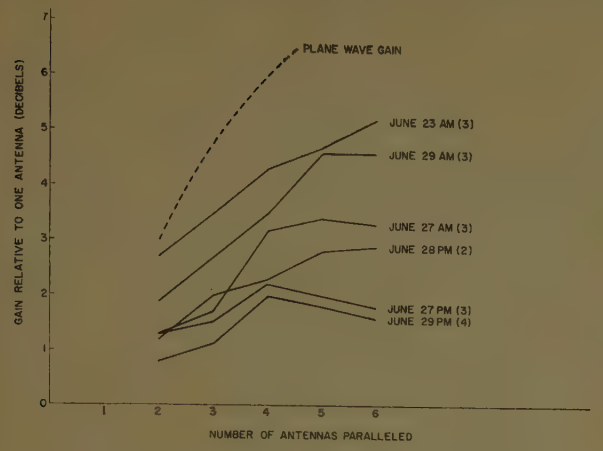


Fig. 3—Some preliminary antenna gain measurements.

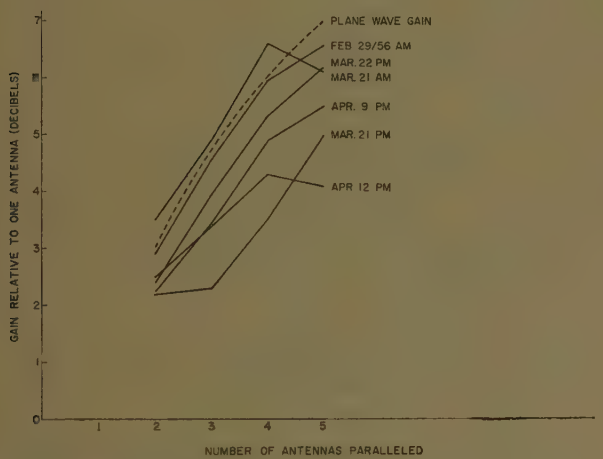


Fig. 4—Some preliminary antenna gain measurements using a reference antenna.

quently a 20-foot aperture was the maximum that could be obtained. As before, the signal received with each antenna arrangement was recorded for three minutes with a time constant of four seconds. The median level was estimated from this recording. The antenna gain relative to that of a single antenna was corrected by the amount of signal variation recorded simultaneously on the reference antenna. Each of these curves is the average of two such observations, the entire experiment taking about 40 minutes to perform.

The large variation in realized gain is again observable. Probably as the result of the use of a reference antenna, this variation is less than that shown in Fig. 3. The observations of February 29 and of the morning of March 21 show realized gains in close agreement with the theoretical plane-wave gain. There is a striking difference in the morning and afternoon observations on March 21.

An Account of the Main Observations

A more elaborate program of experiment and analysis was undertaken following the preliminary observations of the previous section. Both antenna gain and diversity measurements were made using a time constant which was short compared with the fading rate, and a high-speed recorder which could follow the signal fading faithfully.

For the antenna gain measurements, one of the antennas was used as a reference and its signal recorded on one channel of a two-channel Brush recorder, while in turn one, two, three, four, and five antennas were paralleled and the resulting signal recorded on the other channel. A recording of approximately 30 seconds was taken at each antenna configuration, and the entire gain measurement took from 10 to 20 minutes. Cumulative amplitude distributions were determined from all the recordings. The relative antenna gains were then calculated from the average db difference between the distribution curves, between the 20- and 80-percentile levels, and corrected by the signal level change indicated by the reference antenna. From these same records, the fading rate of the signal as a function of antenna aperture was also obtained. The fading rate is defined as "the average number of positive crossings, per second, of the median level."

Diversity measurements were made with the same recorder and recording time. One receiver was connected to an antenna at one end of the array and another connected to each of the other antennas in turn. From these records, the normalized envelope correlation functions $\phi_e(\tau)$ of signals received at points separated by 4, 8, 12, 16, and 20 feet were calculated.

$$\phi_e(\tau) = \frac{\sum [E_1(t_i) - \overline{E}_1][E_2(t_i + \tau) - \overline{E}_2]}{\{\sum [E_1(t_i) - \overline{E}_1]^2 \sum [E_2(t_i + \tau) - \overline{E}_2]^2\}^{1/2}} \quad (1)$$

where $E_1(t_i) - \overline{E}$ is the zero-centered signal envelope at one antenna, and $E_2(t+\tau) - \overline{E}$ is a similar signal at a second antenna a time τ later. For each separation, the correlation function was calculated over a range of time shifts sufficient to include the maximum in the correlation. A set of correlation curves is shown in Fig. 5. The progressive shift in the position of the maximum and the reduction of the maximum value as the separation is increased is typical of the results obtained.

Fig. 6 shows the results of the analysis of nine experiments. In six of the cases, the experimental data and the analysis are relatively complete. In one case (October 30, 1956), the experimental data are complete but the correlation coefficients have not been calculated. In another case (October 10, 1956), no gain measurements were made, and in a third (December 21, 1956), the diversity measurements were not made.

In the upper portions of Fig. 6 are shown the relative gains of the various antenna combinations relative to the gain of a single antenna, and on the same scale the fading rate ($\times 10$) of these antenna combinations relative to the fading rate for a single antenna. Also listed is the signal level S in db relative to the monthly median,

the fading range R being the db difference between the 10 and 90 percentiles, and Z the fading rate for a single antenna. The value of R for a Rayleigh distribution is 13.4 db. In the lower portions of Fig. 8 are shown two correlation coefficient curves, one giving the correlation at zero time displacement as a function of antenna spacing, and the other the maximum correlation coefficient irrespective of time delay for each antenna separation. A third curve gives the time delay t_m at which this maximum occurs. Again, the signal level, the fading range and the fading rate observed during these diversity measurements are noted. At the bottom are listed the time intervals during which the gain and diversity measurements were made. The wind direction and speed, as reported by the Meteorological Service of Canada station at Trenton, about 35 miles south of the midpoint of the path, is also shown. The direction shown in Fig. 8 is relative to the path. A 0° wind is one blowing into the receivers from the direction of the transmitter. There follows a short description of the individual experiments.

April 19, 1956: There is only one point available for the variation of fading rate with antenna aperture. The line, consequently, is shown dashed.

August 17, 1956: The antenna gain measurements were performed in a different manner than in any of the other experiments. The signal was observed for 10 minutes at each antenna configuration and the signal was analyzed on an automatic totalizer.[3] As a result, the gain measurements stretched out over a period of about 70 minutes, and there was a variation of close to 5 db in median signal level from the beginning to the end of the experiment. The signal correlation dropped rapidly with increasing separation and the maximum in the correlation function was not well defined for separations of 12 feet and 16 feet, and did not exist for a 20-foot separation.

October 4, 1956: In this case, the diversity measurements were made in the morning, and the gain measurements in the afternoon. There was an 8-db change in median signal level and a significant change in the fading rate during the four hours separating the two measurements, and it is considered likely that the diversity results are not indicative of the conditions existing during the afternoon.

October 10, 1956: No gain measurements were made and no wind observations are available. At an antenna separation of 20 feet, it became impossible to determine a maximum in the correlation pattern.

October 17, 1956: The signal level was constant during the period of diversity measurements, but dropped 5 db while the gain measurements were being made. It is considered likely that a significant change in the propagation conditions occurred in this short interval.

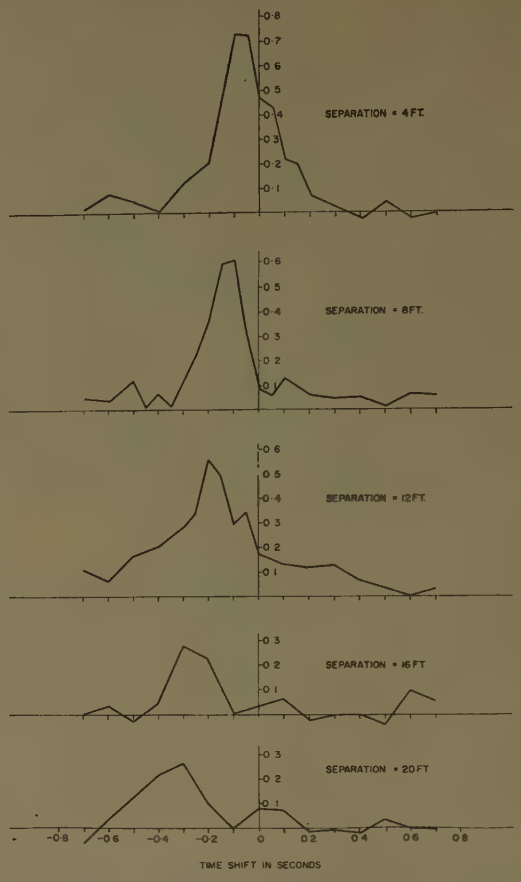


Fig. 5—Cross-correlation functions of signals received on space antennas.

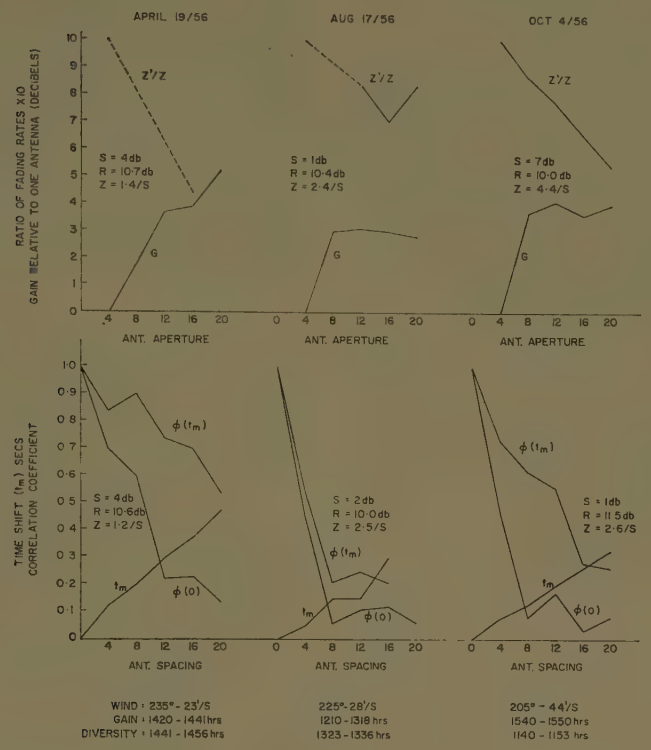
October 26, 1956: Part, at least, of the irregularity in the t_m-vs-separation curve for this date, and for some of the others, is due to the coarseness of the time intervals at which readings were taken. Where the true maximum occurs at a time shift between values for which the correlation coefficient was calculated, it is estimated that the indicated maximum may be low by as much as 0.1.

October 30, 1956: Diversity measurements were made, but the calculations of correlation coefficients have not been made. The curve of time shift in the diffraction pattern has been obtained by taking the average time difference of corresponding maxima in the recorded signals on the spaced antennas. This is possible because of the very good correlation that exists between signals from antennas as far as 20 feet apart.

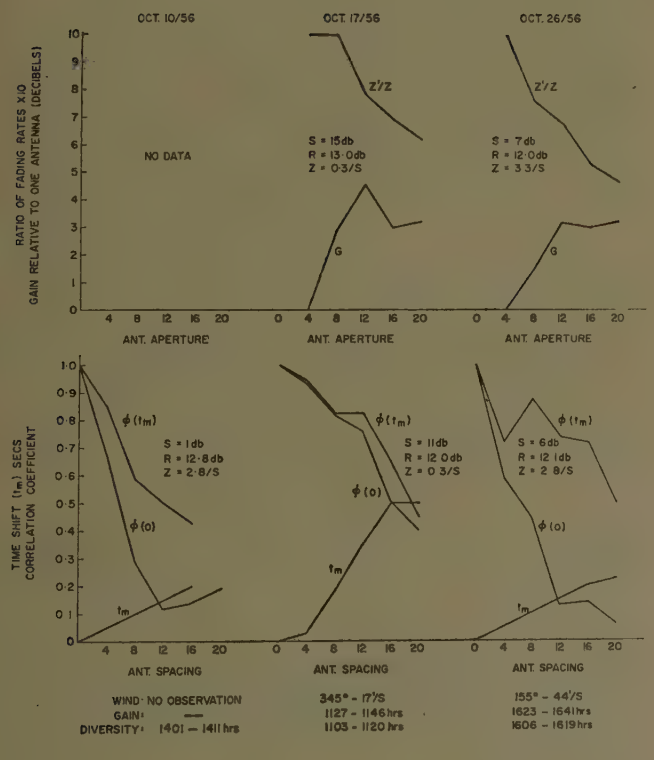
December 21, 1956: No diversity measurements were made.

April 16, 1957: The correlation between signals recorded just 8 feet apart was poor, and consequently

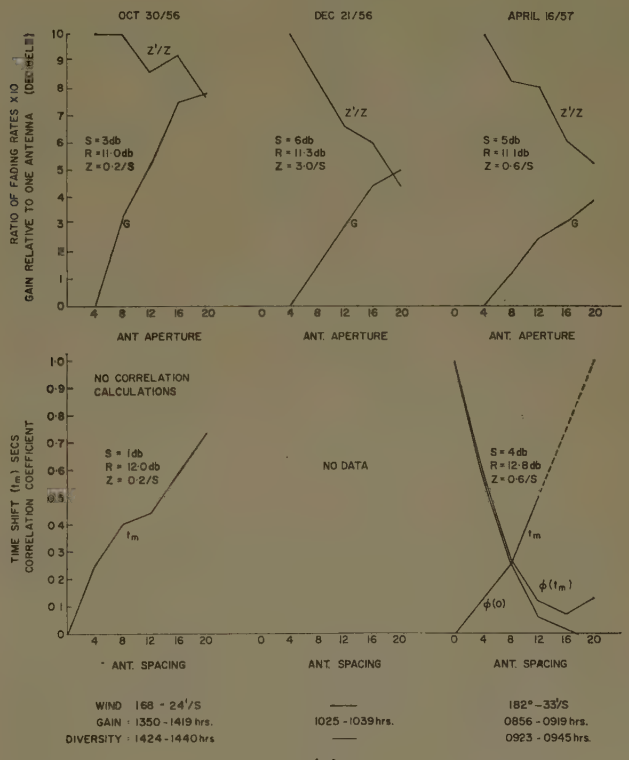
[3] S. A. Stone, "A Totalizer for Obtaining the Amplitude Distribution of a Fading Signal," Natl. Res. Council of Canada, Ottawa, REED Rept. ERB-405; October, 1956.



(a)



(b)



(c)

Fig. 6—Results of antenna gain and diversity measurements. (See text for definitions of Z, Z', G, $\phi(t_m)$, $\phi(0)$, t_m.)

the time-shift curve is not very reliable and could not be read at an antenna separation of 16 feet.

Realized Antenna Gain

The theories of tropospheric scattering have, in the main, called for an aperture-to-medium coupling which is essentially independent of meteorological conditions. Waterman[4] gives the following formula for the present situation of one broad-beam antenna and the other antenna narrow in azimuth only:

$$L_c = \sqrt{5^{2/m} - 1}\,\frac{d/a}{\phi} \tag{2}$$

where

L_c = the aperture-to-medium coupling loss,
d = the distance between transmitter and receiver,
a = the modified earth's radius,
ϕ = the narrow azimuthal beamwidth,
m = the exponent for the angular dependence of the scattering cross section σ,

$$\sigma(\theta) = \frac{l}{\sin^m(\theta/2)}, \text{ and}$$

θ = the scattering angle.

The exponent m has different values for different scattering models. For the Booker-Gordon mode,[5,6] $m=4$. For the models proposed by Villars and Weisskopf,[7] and by Wheelon,[8] the exponent has a value of 5. A value of $m=11/3$ is obtained by both Batchelor[9] and Silverman.[10] For each model, however, meteorological conditions enter only through the modified earth's radius factor. It is unlikely that significant changes in a can occur, at least for the link considered here, without producing a distinguishable superrefracted signal.

Significant changes in the aperture-to-medium coupling loss are, however, observed in the results contained in this paper. Fig. 3 shows some indication of a diurnal pattern in the measurement of aperture-to-medium coupling loss. Fig. 4 shows two instances where the realized and plane-wave gains are approximately equal. It appears, therefore, that a model is required which permits a variation in the scattering cross section's dependence on scattering angle.

Before considering further the antenna gain measurements included in Fig. 6, it is worthwhile to relate these gain measurements to the measured correlation coefficients. The measured correlations were of the envelopes of the received signal, whereas the RF correlation is required for a calculation of the expected gain. If $\phi(kl)$ is the normalized RF correlation (zero time displacement) between signals observed a distance kl apart, then it is easily shown that a gain of n similar antennas in parallel relative to the gain of a single antenna is,

$$G(n) = 1 + \sum_{k=1}^{n-1} \frac{2(n-k)}{n}\phi(kl) \tag{3}$$

where l is the distance between adjacent antennas.

The correlation function ϕ has the form,[11]

$$\phi(x) = \sigma(x)\cos[\lambda(x)] \tag{4}$$

where $\sigma(x)$ is related to the measured envelope correlation function $\phi_e(x)$ as follows,

$$\phi_e(x) = \frac{\pi}{4-\pi}\left(\frac{1}{4}\sigma^2 + \frac{1}{64}\sigma^4 \cdots\right). \tag{5}$$

Thus, $\sigma(x)$ can be derived from the measured envelope correlation, but the term $\lambda(x)$, due to the phase fluctuations, cannot be so derived. Although the gain of the parallel antennas cannot be calculated from the measured correlation coefficients, it was worthwhile to check the internal consistency of the measurements by a calculation assuming $\phi(x)=\phi_e(x)$. The results are shown in Fig. 7. The points fall within a reasonably narrow band except for the October 4 results, and two points of both the August 17 and October 17 results. In each of these cases there is good cause for believing, as already stated, that the results of the diversity measurements were not representative of conditions during the gain measurements. It appears, therefore, that the gain measurements may be considered with some confidence.

The variability in the aperture-to-medium coupling loss is again observed in the eight gain curves plotted in Fig. 6. In particular, the results for October 30 show no aperture-to-medium coupling loss even for an aperture of 20 feet. This would not be an unexpected result if the signal received were due to superrefraction. However, the signal level is approximately equal to the monthly median and the experience with this path[12] is that there is no evidence of superrefraction even at the 10-percentile level. In addition, the signal is approximately Rayleigh-distributed and the fading rate, while low, will be shown to be a result of the mean wind being directed very closely along the path. Thus, except for the gain

[4] A. T. Waterman, Jr., "Some generalized scattering relationships in trans-horizon propagation," Proc. IRE, vol. 46, pp. 1842–1848; November, 1958.
[5] H. G. Booker and W. E. Gordon, "A theory of radio scattering in the troposphere," Proc. IRE, vol. 38, pp. 401–412; April, 1950.
[6] W. E. Gordon, "Radio scattering in the troposphere," Proc. IRE, vol. 43, pp. 23–28; January, 1955.
[7] F. Villars and V. F. Weisskopf, "On the scattering of radio waves by turbulent fluctuations of the atmosphere," Proc. IRE, vol. 43, pp. 1232–1239; October, 1955.
[8] A. D. Wheelon, "Spectrum of turbulent fluctuations produced by convective mixing of gradients," *Phys. Rev.*, vol. 105, pp. 1706–1710; March 15, 1957.
[9] A. K. Batchelor, "The Scattering of Radio Waves in the Atmosphere by Turbulent Fluctuations in Refractive Index," Cornell Univ., Ithaca, N. Y., Res. Rept. EE 262; September 15, 1955.
[10] R. A. Silverman, "Turbulent mixing theory applied to radio scattering," *J. Appl. Phys.*, vol. 27, pp. 699–705; July, 1956.
[11] R. Price, "A note on the envelope and phase-modulated components of narrow-band Gaussian noise," IRE Trans. on Information Theory, vol. IT-1, pp. 9–13; September, 1955.
[12] Doherty and Neal, *op. cit.* The limited occurrence of superrefraction was confined to the summer months.

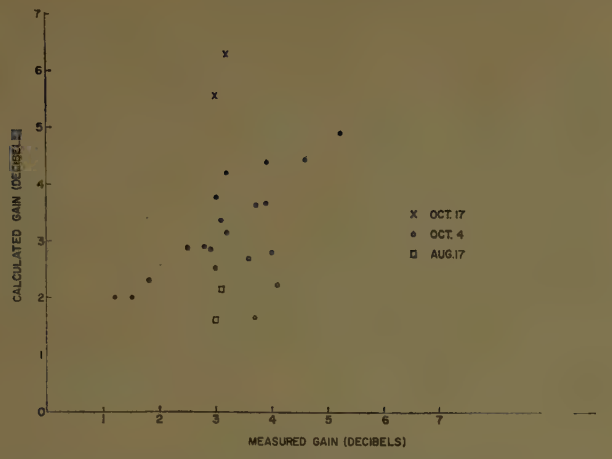


Fig. 7—Comparison of measured gains with gains calculated from results of diversity measurements.

measurement itself, there is no reason to believe that this signal differs from the normal scatter signal on this path.

The evidence presented in Figs. 3 and 4, and again in Fig. 6, suggests very strongly that there is a basic variability in the scattering mechanism. The data of Fig. 6 in particular are sufficiently accurate so that the variability in the results cannot be assigned to experimental error. In addition, the strength of the signal, the amplitude distribution, and the relationships (which are discussed in the next section) between the fading rate, the correlation distance and the mean wind, are all indicative of a tropospheric scatter signal and not of another propagation mechanism. Bolgiano[13] has found a similar variability in the wavelength dependence and has been led to produce a model in which the m of (2) varies with the dynamic stability of the atmosphere. No attempt has yet been made to relate the measurements presented here with the meteorological situation in the manner suggested by Bolgiano's work.

Anisotropy in the turbulence is a consequence of Bolgiano's model, in which atmospheric stability inhibits vertical motion. The effect of anisotropic turbulence is treated by Staras[14] in terms of a parameter r which is the ratio between vertical and horizontal scales of turbulence. In order to account for the results of Fig. 6, r must vary between unity and approximately $\frac{1}{8}$.

Fading and the Wind

Following Gordon[15] and Rice,[16] we assume a space correlation function for the envelope of the signal in the vicinity of the receiver of the form,

$$\phi_e(x, y, z) = \exp - \{(x/x_0)^2 + (y/y_0)^2 + (z/z_0)^2\}, \quad (6)$$

where x, y, and z are measured respectively, horizontally across, horizontally along the path, and vertically. x_0, y_0, and z_0 are the correlation or diversity distances defined by

$$\phi_e(x_0, 0, 0) = \phi_e(0, y_0, 0) = \phi_e(0, 0, z_0) = e^{-1}.$$

As a result of drift of the scatterers with the mean wind in the scattering volume, the pattern of signal strength in the vicinity of the receiver will move in a related manner. The correlation function, including time as a variable, will therefore have the form,

$$\phi_e(x, y, z, t) = \exp - \left\{\left(\frac{x - v_x t}{x_0}\right)^2 + \left(\frac{y - v_y t}{y_0}\right)^2 + \left(\frac{z - v_z t}{z_0}\right)^2 + \left(\frac{t}{t_0}\right)^2\right\}, \quad (7)$$

where v_x, v_y, v_z, are the x, y, and z components of the velocity of the diffraction pattern at the receiving site; and t_0 is the period or time constant of the turbulence.

Implicit in (7) is the assumption that in the absence of turbulent motions, the diffraction pattern will not change in time when viewed in a co-ordinate system moving with the velocity v_x, v_y, v_z. An examination of the phase relationships in such a moving system shows that this is a good approximation for motion normal to the path. For motions along the path where the term in y is dominant, the approximation is poorer. Sample calculations indicate that when $v_y t > 2y_0$, the changing phase relationships will begin to change the diffraction pattern even if there is no turbulence. In the analysis which follows, this failure of (7) is of no importance except in the section on turbulent velocities where this point is again discussed briefly.

For the diversity measurements described in this paper $y = z = 0$, and if, in addition, we assume $v_z = 0$, (7) reduces to

$$\phi_e(x, t) = \exp - \left\{\left(\frac{x - v_x t}{x_0}\right)^2 + \left(\frac{v_y t}{y_0}\right)^2 + \left(\frac{t}{t_0}\right)^2\right\}. \quad (8)$$

If the scattering occurs in the neighborhood of the midpoint of the path, the relationship, to a good approximation, between the diffraction pattern drift velocity and the mean wind V in the scattering volume is

$$v = 2V. \quad (9)$$

A partial check on the validity of (8) is provided by Fig. 8. The correlation coefficients for zero time shift have been plotted against the normalized space separation x/x_0. Values from all the diversity measurements of Fig. 6 are included in Fig. 8. A curve of the form $\exp(-x^2)$ has been superimposed for comparison, and is seen to agree well with the experimental points for correlation coefficients greater than 0.2.

[13] R. Bolgiano, Jr., "A Meteorological Interpretation of Wavelength Dependence in Transhorizon Propagation," Cornell Univ., Ithaca, N. Y., Res. Rept. EE 385; September, 1958.

[14] H. Staras, "Antenna-to-medium coupling loss," IRE Trans. on Antennas and Propagation, vol. AP-5, pp. 228–231; April, 1957.

[15] W. E. Gordon, "The Scattering of Radio Waves by Turbulence in the Troposphere," Cornell Univ., Ithaca, N. Y., Res. Rept. EE 163; September, 1953.

[16] S. O. Rice, "Statistical fluctuations of radio field strength far beyond the horizon," Proc. IRE, vol. 41, pp. 274–281; February, 1953.

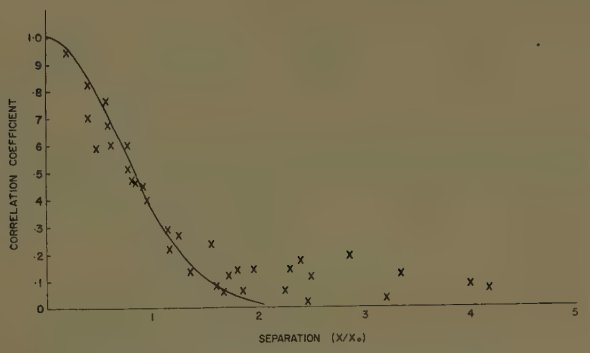


Fig. 8—Signal correlation normal to the path (solid curve is exp $-(x/x_0)^2$).

Calculation of the Mean Wind

If the cross-correlation function of the signals received on spaced antennas is calculated, the maximum in the correlation function is found to be shifted in time, as in Fig. 5. By setting $\partial\phi_e/\partial t = 0$, the time displacement of the maximum is found to be

$$t_m = \frac{xv_x/x_0^2}{\frac{v_x^2}{x_0^2} + \frac{v_y^2}{y_0^2} + \frac{1}{t_0^2}} . \tag{10}$$

The linear relationship between t_m and x is to be observed in Fig. 6. The value of the correlation function at the maximum is,

$$\phi_e(x, t_m) = \exp - \left\{ \frac{x^2}{x_0^2} - \frac{xv_x t_m}{x_0^2} \right\} \tag{11}$$

where

$$v_x = \frac{x_0^2}{xt_m} \ln \phi_e(x, t_m) + \frac{x}{t_m}$$

$$= \frac{x}{t_m} - \frac{x_0^2}{xt_m} \left| \ln \phi_e(x, t_m) \right| . \tag{12}$$

The first term has been called the "apparent velocity."[17]

This expression has been used to calculate the component of the mean wind normal to the path. In this calculation, a straight line has been fitted to the t_m-vs-x plots of Fig. 6. In most cases, this could be done without difficulty. Using the fact that t_m is proportional to x, it can be seen from (11) that $\ln \phi_e(k, t_m) \propto x^2$. Consequently, $\ln \phi_e$ was plotted against x^2 and an attempt was made to fit a straight line to the results. In three of the six cases considered, this could be done with reasonable accuracy. In two of the three remaining cases (April 16, 1957; August 17, 1956), the correlation had dropped to a low value with an 8-foot separation. It is believed that the statistical fluctuation in the low values of correlation could account, in part, for the nonlinearity in these plots. A similar effect may be observed in Fig. 8.

[17] S. N. Mitra, "A radio method of measuring winds in the ionosphere," *Proc. IEE*, vol. 96, pt. III, pp. 441–446: 1949.

Table II gives the values of the normal component of the mean wind as computed, using (12) and (9). A description of the fit of the ln $\phi(x, t_m)$-vs-x^2 to a straight line is included. The measured winds were obtained from the Meteorological Service of Canada's observing station at Trenton, Ontario, about 35 miles south of the midpoint of the path; they are for an altitude of 5000 feet, the approximate height of the intersection of the horizon lines from the transmitter and receiver. The apparent velocity given by x/t_m is also listed. With the exception of the October 26, 1956 data, the agreement between the computed and mesaured winds is good, and tends to be better when there is good linearity in the ln ϕ-vs-x curve. The measured wind for the October 26 data was extrapolated from measurements made from the surface to 4000 feet, at which height the observations were terminated. A strong wind shear was evident in the observations over the first 4000 feet, and a front was in the general vicinity of the path. It is entirely possible, therefore, that the measured wind quoted in Table II is in error. A 30° shift in wind direction without change in wind speed would be sufficient to make the measured and calculated winds equal.

TABLE II

Component of the Mean Wind Normal to the Path (Feet/Second)

Date	Linearity ln ϕ vs x^2	$\frac{1}{2}\frac{x}{t_m}$	V_x (feet/second)	
			Computed	Measured
April 19, 1956	good	20	17	19
August 17, 1956	poor	31	13	20
October 4, 1956	poor	60	26	19
October 17, 1956	good	18	4	4
October 26, 1956	good	43	38	19
April 16, 1957	fair	14	2	1

Fading Rate

The observed fading rate of a signal received on a small antenna (aperture less than the diversity distance) is of the order of the reciprocal of the correlation time. From Rice[18] the average number of positive crossings of the median per second is, approximately

$$Z = -\frac{1}{2\pi} \frac{\partial^2}{\partial t^2} [\phi_e(0, 0)]$$

$$= \frac{1}{\pi} \left\{ \frac{v_x^2}{x_0^2} + \frac{v_y^2}{y_0^2} + \frac{1}{t_0^2} \right\}^{1/2} . \tag{13}$$

It can be seen from (13) that attempts to correlate the fading rate with the component of wind normal to the path need not be very successful. In fact, correlation coefficients as low as 0.45 have been observed.[19]

[18] S. O. Rice, "Statistical properties of a sine wave plus random noise," *Bell Sys. Tech. J.*, vol. 27, pp. 109–157; January, 1948.

[19] B. C. Angell, J. B. L. Foot, W. J. Lucas, and C. T. Thompson, "Propagation measurements at 3480 mc/s over a 173-mile path," *Proc. IEE*, pt. B, suppl. no. 8, pp. 128–142; 1958.

From (13) and (10), an expression for the fading rate of the signal received on a small antenna may be derived,

$$Z = \frac{1}{\pi}\left[\frac{xV_x}{t_m x_0^2}\right]^{1/2}. \tag{14}$$

Fading rates have been calculated for the six cases considered in the previous section. The fading rates have been calculated using both the calculated and measured values of the normal component of the mean wind. The results are shown in Table III.

TABLE III

FADING RATE OF SIGNAL RECEIVED ON SMALL ANTENNA

Date	Fading Rate		
	Calculated from		Measured
	Calculated v_x	Measured v_x	
April 19, 1956	1.1	1.3	1.2
August 17, 1956	2.6	3.6	2.5
October 4, 1956	3.5	2.6	2.6
October 17, 1956	0.3	0.6	0.3
October 26, 1956	2.9	1.5	2.8
April 16, 1957	0.5	0.4	0.6

With the exception of the October 4 data, the comparison between the measured fading rates and those calculated using the calculated values of the wind v_x is very good. It is worthy of note that the October 26 data are at least selfconsistent. Thus, although there was a ratio of almost two-to-one between the calculated and measured wind speeds, the fading rate corresponds to that calculated from the calculated wind speeds.

Eq. (13) is valid when the antenna aperture is less than x_0. For an antenna whose aperture is large in the x-direction, the term $(v_x/x_0)^2$ must be modified to allow for the averaging that occurs over the aperture. The effective correlation distance is of the order of the antenna dimensions and the expression for the fading rate takes the form

$$Z' = \frac{1}{\pi}\left\{\frac{v_x^2}{L^2} + \frac{v_y^2}{y_0^2} + \frac{1}{t_0^2}\right\}^{1/2} \tag{15}$$

where L is the aperture dimension in the x-direction.

From (10) and (12), it can be shown that

$$\frac{v_y^2}{y_0^2} + \frac{1}{t_0^2} = \frac{v_x}{x t_m} \ln \phi(x, t_m) \tag{16}$$

whence,

$$Z' = \frac{1}{\pi}\left\{\frac{v_x^2}{L^2} + \frac{v_x}{x t_m} \ln \phi(x, t_m)\right\}. \tag{17}$$

The fading rates have been calculated using (17). The results are presented in Table IV in terms of the ratio of the fading rate observed with an aperture L to that observed with an aperture of 4 feet.

TABLE IV

FADING RATE FOR VARIOUS ANTENNA APERTURES (FADING RATE RELATIVE TO THAT FOR SMALL APERTURE)

L	April 19, 1956		August 17, 1956		October 4, 1956	
	Calculated	Measured	Calculated	Measured	Calculated	Measured
8	1.00		0.86		0.68	0.87
12	0.91		0.81	0.83	0.53	0.78
16	0.76	0.44	0.79	0.71	0.46	0.66
20	0.64		0.78	0.83	0.43	0.54
	October 17, 1956		October 26, 1956		April 16, 1957	
	Calculated	Measured	Calculated	Measured	Calculated	Measured
8	1.00	1.00	1.00	0.76	0.98	0.83
12	1.00	0.78	0.77	0.68	0.96	0.81
16	1.00	0.69	0.62	0.53	0.97	0.61
20	1.00	0.62	0.54	0.47	0.95	0.53

The agreement between the calculated and measured fading rates is not good. It has already been noted in connection with the gain measurements that the diversity measurements of October 17, 1956, on which the calculated fading rates are based, are probably not indicative of conditions during the period when the antenna aperture was varied. The data of August 17 and October 26 show fair agreement between the calculated and measured fading rates, but those for October 4, and particularly for the following April 16, do not. In the latter case in particular, it appears certain that some factor affecting the contribution to the fading rate from the wind component v_y has not been taken into account.

Turbulent Velocity

On the basis of the analysis presented in this paper, the term $(v_y^2/y_0^2)+(1/t_0^2)$ is given by (16). It is not possible to derive a value for y_0 from the experiments presented here. Gordon gives the following relationship between x_0 and y_0:[20]

$$y_0 = x_0^2/2\lambda. \tag{18}$$

Using (16) and (18), an estimate of t_0 may be made. The results are shown in Table V.

TABLE V

CALCULATED PERIOD OR TIME CONSTANT OF ATMOSPHERIC TURBULENCE

Date	t_0 (seconds)
April 19, 1956	0.7
August 17, 1956	0.2
October 4, 1956	0.3
October 17, 1956	1.3
October 26, 1956	0.3
April 16, 1957	1.2

[20] W. E. Gordon, *op. cit.* footnote 15, p. 52.

The periods listed in Table V are to be associated with an eddy scale l.

$$l = \frac{2\pi}{K} = \frac{1}{\frac{2\pi}{\lambda}\sin\frac{\theta}{2}}$$

$$= 2.7 \text{ meters.}$$

The values for t_0 must of course be treated with a great deal of caution. As mentioned earlier, some reshuffling of the diffraction pattern occurs even in the absence of turbulent motions. In this analysis, the result of this effect is included in t_0. The t_0 of Table V must therefore be considered as an upper limit to the period of the turbulence. An estimate of the correction to be applied could be made, but it is not considered worthwhile in view of the lack of a measured value of y. An additional factor, which might possibly be important in some cases, would be the presence of a vertical air current. The effect of any such vertical air movements is also included in the t_0 of table V.

ACKNOWLEDGMENT

G. Neal of these Laboratories took part in all the experimental work described here and his help is gratefully acknowledged. Mrs. E. Peters and Mrs. J. Burrell did much of the data reduction and all the computations. Various members of the Meterological Service of Canada were extremely helpful, and particular thanks are due to the personnel of the Weather Office at the Trenton Air Station.

Sweep-Frequency Studies in Beyond-the-Horizon Propagation*

W. H. KUMMER†

Summary—**This paper considers the bandwidth characteristics of the propagating medium in tropospheric beyond-the-horizon propagation.**

To study this problem, a frequency-sweep experiment was performed over a 171-mile experimental circuit. A 4.11-kmc transmitter was frequency modulated at a 1000-cps rate over a 20-mc band. The receiver was swept nonsynchronously over the same band at a 30-cps rate. The resultant pulses were displayed on an oscillograph and photographed at the rate of one frame every two seconds.

The experiment used a 28-foot transmitting antenna and 8-, 28- and 60-foot receiving antennas.

Sequences of selected sweep-frequency pictures are shown for various antenna combinations and transmission conditions. The bandwidths from the experiment are compared with a calculation based on the common volume geometry.

Photographs of signals received simultaneously from a twin-feed horizontal diversity system are also shown and discussed.

INTRODUCTION

COMMERCIAL and military tropospheric beyond-the-horizon systems are used to transmit voice and television, as well as data.[1] The complexity of these systems makes it difficult to use them in the measurement of the instantaneous bandwidth.

In the experiments described in this paper, the instantaneous bandwidth is measured by a frequency-sweep method. It is possible to obtain equivalent information using pulse techniques[2] but the interpretation of the results is rather difficult because of pulse distortion and overlapping.

The experimental circuit used is 171 miles in length between Pharsalia, N. Y., and Crawford Hill, N. J. The terrain between locations is mountainous. There is no obstruction at either terminal.

The bandwidth can be estimated by calculating the relative delay from the common volume geometry. This calculation of bandwidth is based on the maximum delays produced by the path differences defined by lines drawn through the 3-db points of the antenna patterns. From the common volume concept it is apparent that the delays will be greatest in the vertical plane. The 3-db points of the antenna patterns are 0.3°, 0.6° and 2.2° for the 60-, 28-, and 8-foot paraboloids, respectively, as shown in Fig. 1. This method neglects the fact that the atmospheric discontinuities responsible for the propagation are weaker at higher elevations, and the reflection coefficient of these discontinuities decreases with larger grazing angles. These two effects decrease

* Manuscript received by the PGAP, February 9, 1959; revised manuscript received, July 23, 1959. This paper was presented at URSI Meeting, University Park, Pa., October 22, 1958.

† Microwave Laboratory, Hughes Aircraft Co., Culver City, Calif.; formerly at Bell Telephone Labs., Holmdel, N. J.

[1] PROC. IRE, vol. 43; October, 1955.

[2] J. H. Chisholm, P. A. Portmann, J. T. deBettencourt and J. F. Roche, "Investigations of angular scattering and multipath properties of tropospheric propagation of short radio waves beyond the horizon," PROC. IRE, vol. 43, pp. 1317–1335; October, 1955.

the power which is received from higher elevations; if they are taken into account, it is found that the atmosphere limits the effective beamwidth of wide beam antennas. From beam-swinging experiments[3] we have found that the atmosphere effectively limits the beamwidth in the vertical plane to about 1° at 4 kmc.

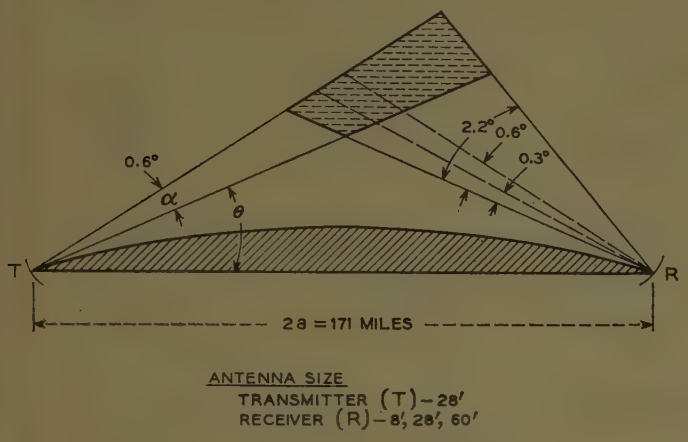


Fig. 1—Vertical beamwidth of antennas.

With the aid of Fig. 1 we see that the path length difference, δ, will be

$$\delta = a\theta^2\left[\left(1+\frac{a_T}{\theta}\right)\left(1+\frac{a_R}{\theta}\right)-1\right]$$

where a_T and a_R are the 3-db beamwidths in radians of the transmitting and receiving antennas; θ is the angle in radians between the lower edge of the beam at the horizon and the straight line joining the terminals. For our path, $2a = 171$ miles and $\theta = 0.9°$ (4/3 earth).

Using the parameters for this circuit we get the bandwidths shown in Table I.

TABLE I

Antenna Sizes Xm.	Rec.	Beamwidths (3 db) Xm. a_T	Rec. a_R	Path Length Difference (feet)	Bandwidth in mc
28′	60′	0.6°	0.3°	136	7.4
28′	28′	0.6°	0.6°	198	5.1
28′	8′	0.6°	2.2°*	281	3.6

* Use only 1° due to atmosphere.

We define the bandwidth as the frequency difference between two adjacent amplitude minima.

Depending on the structure of the atmosphere, these bandwidths will change. The actual bandwidth will be determined by the distribution and strength of the discontinuities in the common volume region. For example, if there were only one discontinuity, then one would expect an infinite bandwidth.

[3] W. H. Kummer and D. C. Hogg, "Characteristics of Signals Received on a Large Aperture Antenna in Propagation Beyond the Horizon," presented at URSI Meeting, Washington, D. C.; May 25, 1957.

Experimental Set-up

The transmitter consists of a klystron oscillator frequency-modulated by a 1000-cps sawtooth voltage applied to the repeller. This signal is amplified by a stagger tuned klystron-amplifier. The output power is about 10 watts with a bandwidth of 10 mc at the 1-db points and 15 mc at the 3-db points. No ripples occur across the band.

The receiver consists of a triple detection set with two IF amplifiers whose center frequencies are 66 mc and 3 mc. The output of the second IF amplifier (150-kc bandwidth) is connected to the Y-axis of an oscilloscope (Fig. 2). The first beating oscillator is frequency-modulated at a 30- or 60-cps rate in a manner similar to the transmitter. These rates were chosen to be an order of magnitude higher than any fading rate which might be encountered on this circuit.

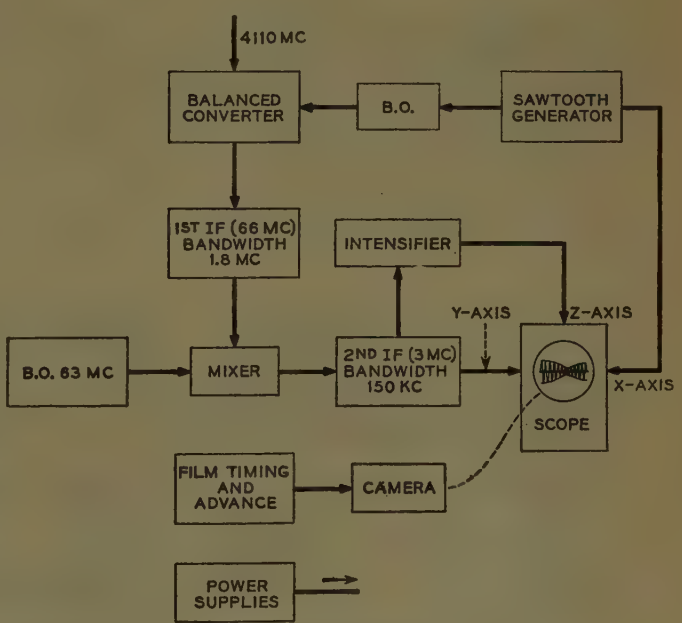


Fig. 2—Sweep receiver.

With the aid of Fig. 3 we see that whenever the difference between the transmitter frequency and the beating oscillator frequency is 66 mc, a pulse will appear on the oscilloscope connected to the last IF amplifier. The oscilloscope presentation is amplitude vs frequency.

Results

A series of photographs taken on different days with the frequency sweep equipment are shown in Figs. 4–8. Each run consisted of a series of 150 frames taken at $2\frac{1}{2}$ second intervals with a 1/10 second exposure. Selected sequences from several runs are shown. In the first three sets, Figs. 4, 5, and 6, the antenna sizes (28′–60′) remain fixed. These sets were taken on different days to show the variability of transmission conditions. Fig. 4 shows

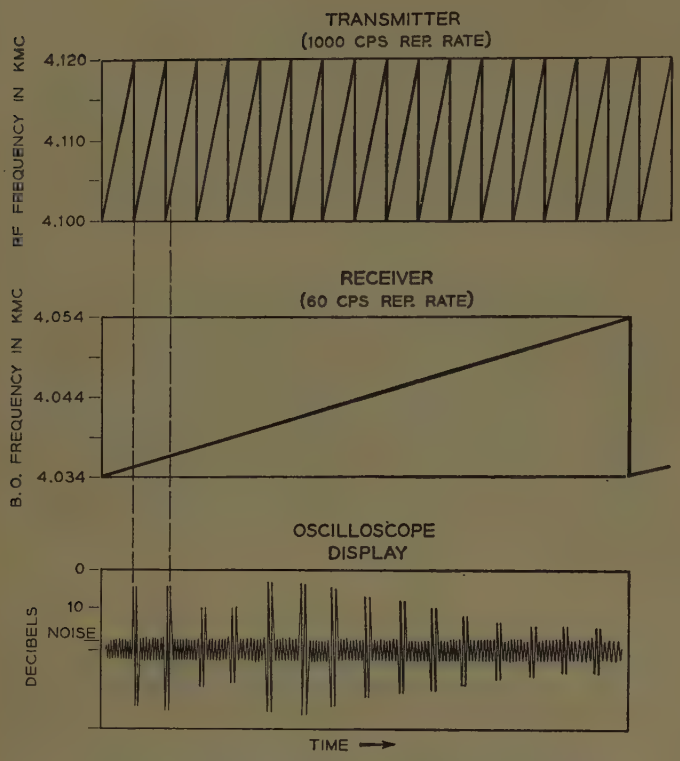


Fig. 3—Combination of transmitter and receiver sweep-frequency waveforms.

the photographs for the 28′–60′ antenna combinations on a day when the median signal level was high with a very slow fading rate (for a 4-kmc-cw signal). Since the fading rate is slow, the sequence appears continuous. The signal goes through one deep fade for the set shown. It should be noted that the amplitude fall-off at the extremities of the photographs is due to the limited bandwidth of the transmitter. In this case the usable bandwidth of the medium is over 15 mc. The wide bandwidth suggests that the rather strong discontinuities primarily responsible for the propagation are contained in a rather small height interval.

In Fig. 5 we observe at times a narrower bandwidth [Fig. 5(h)] as well as a very wide one [Fig. 5(o)]. In this set, the bandwidth is closer to what we predicted from our simple calculation. The characteristic which is most noticeable is the variability from frame to frame. Had a continuous record been taken, continuous changes would have been noted.

Fig. 6 shows, on the average, a broader bandwidth than the previous one. However, the narrowest one [Fig. 6(b)] is about 6 mc wide. Fig. 6 is for the 28′–60′ antenna combination, as were all the preceding ones. Fig. 6, and the ones to follow were taken on the same day, with the receiving antenna as the variable.

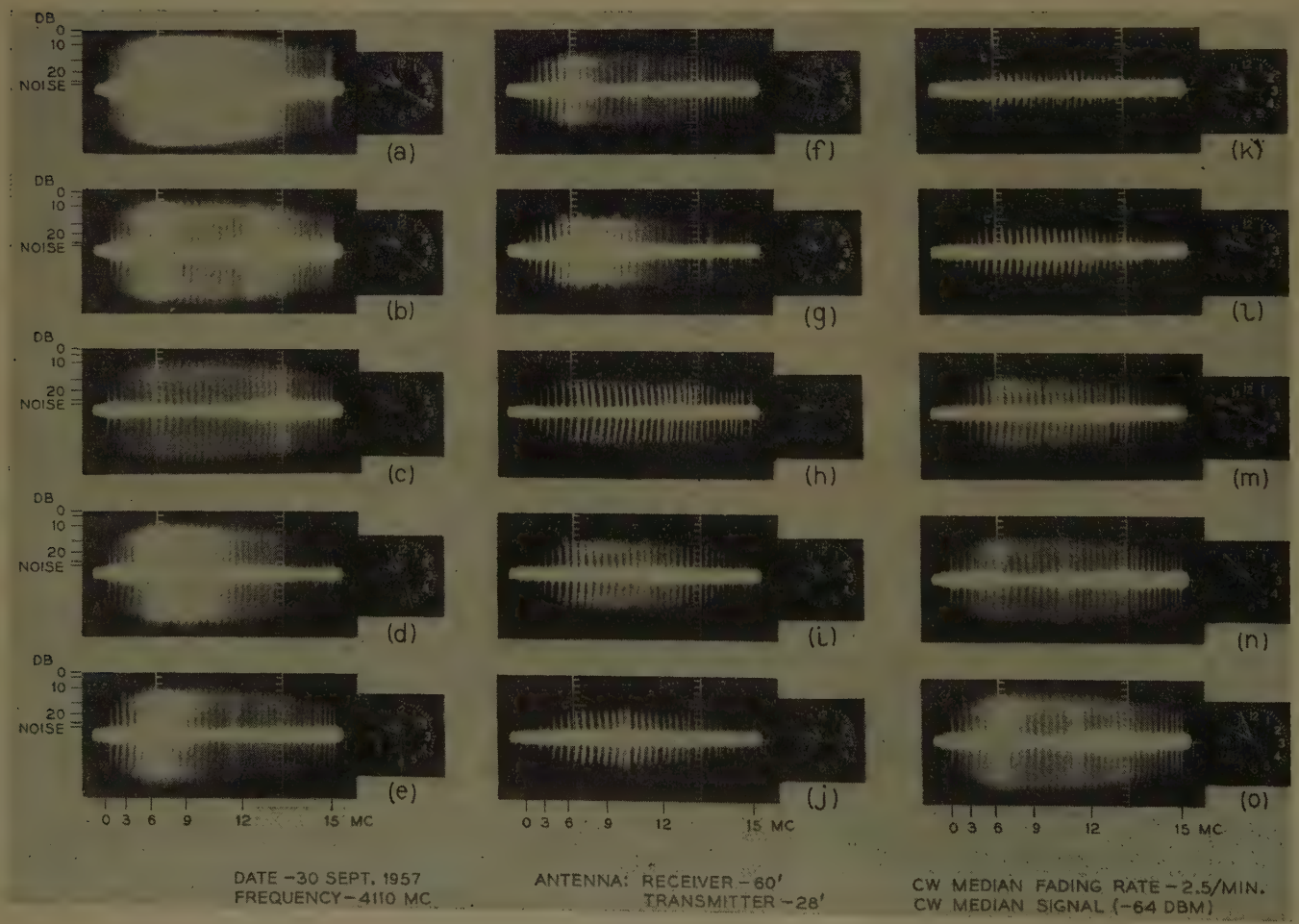


Fig. 4—Sweep-frequency photographs.

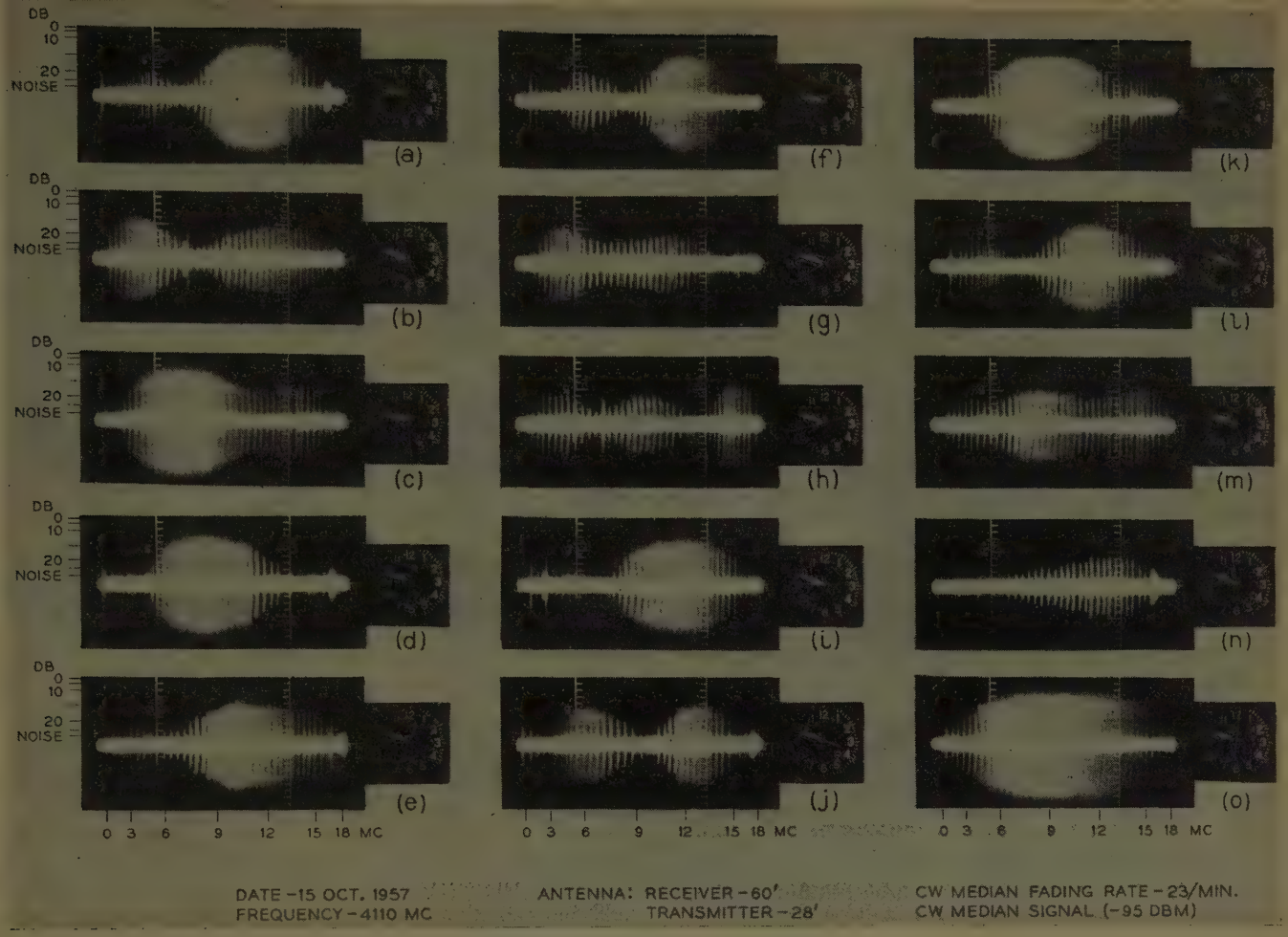


Fig. 5—Sweep-frequency photographs.

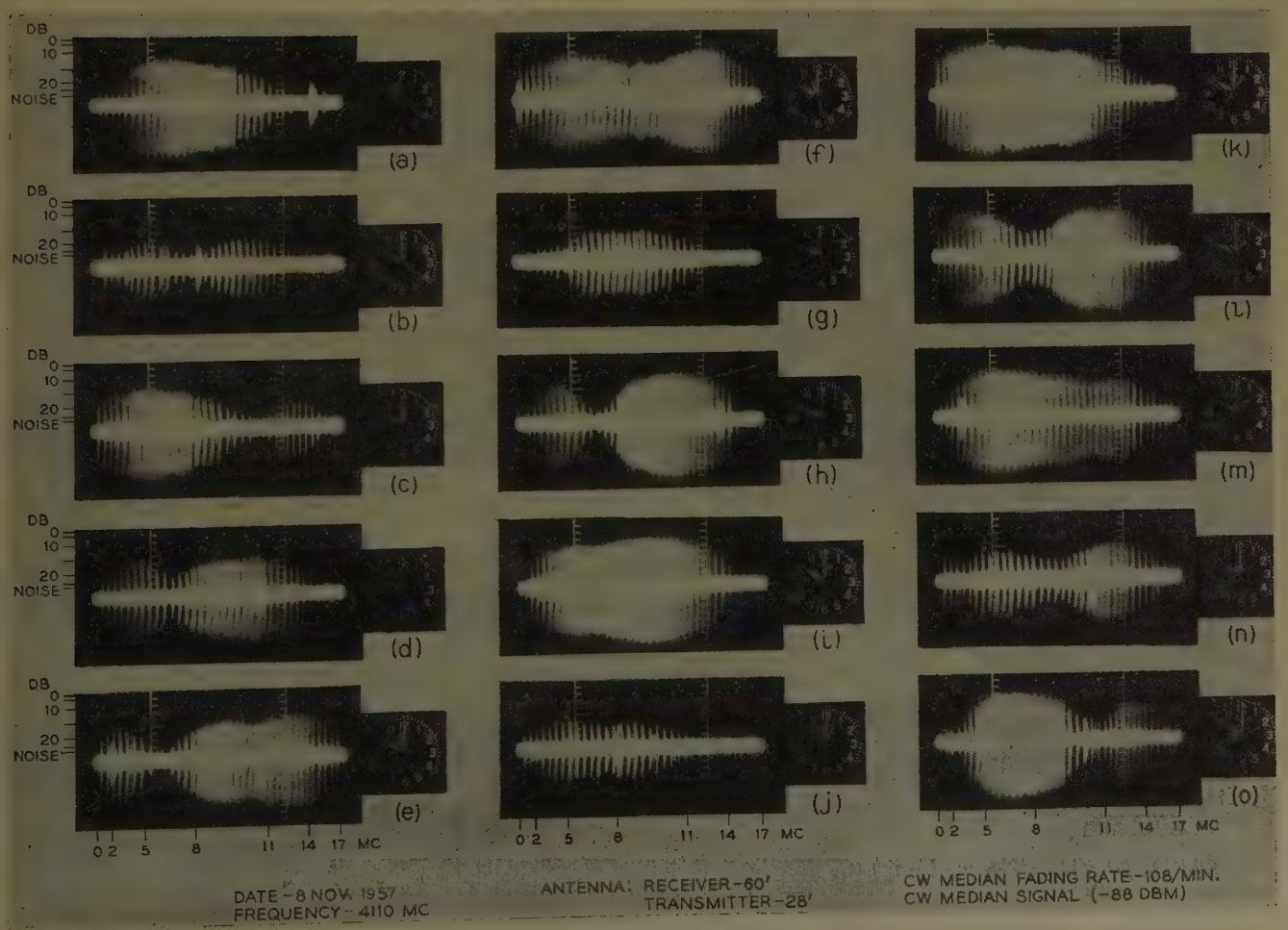


Fig. 6—Sweep-frequency photographs.

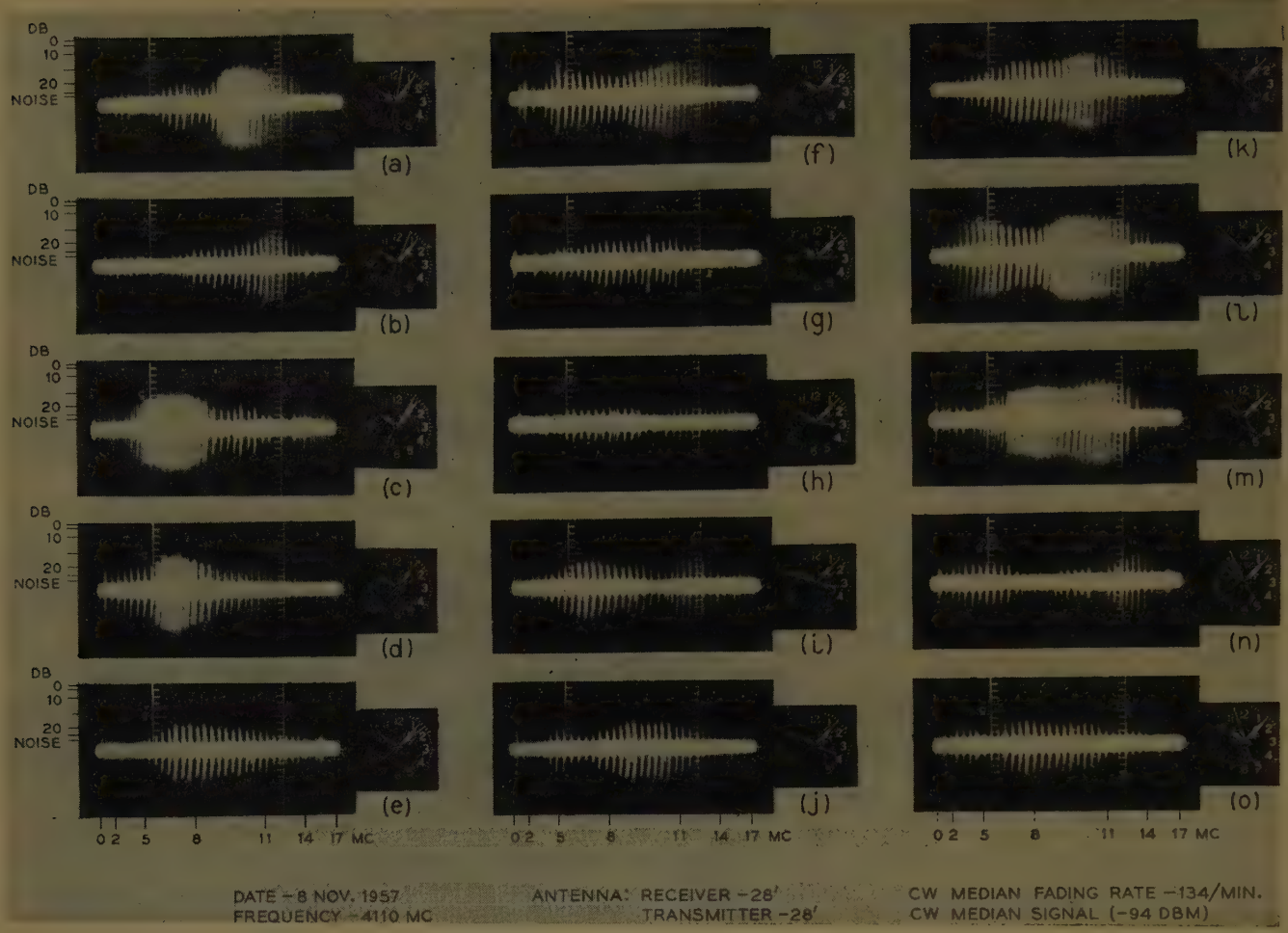


Fig. 7—Sweep-frequency photographs.

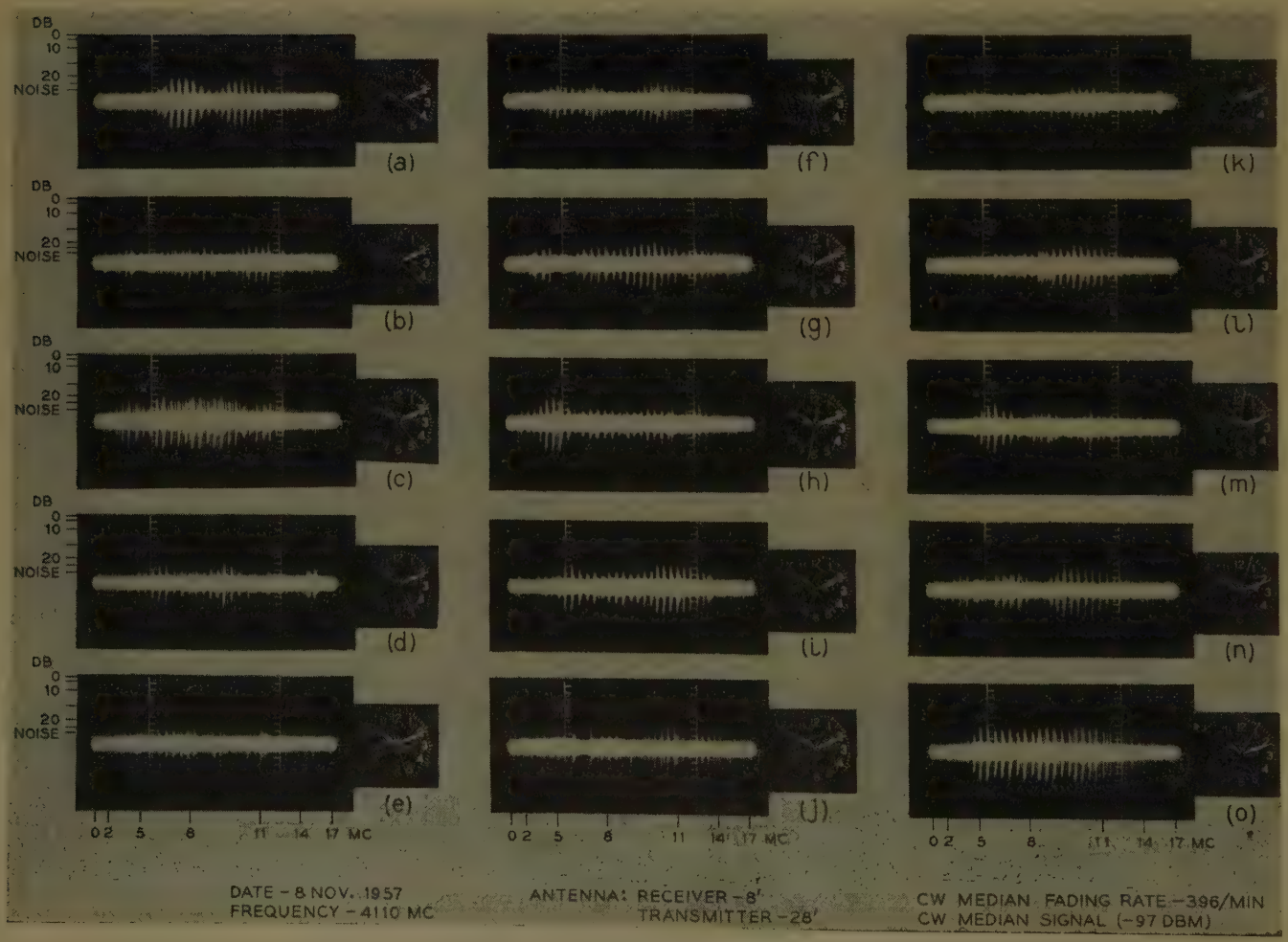


Fig. 8—Sweep-frequency photographs.

Fig. 7 shows the received signal for the 28′–28′ antennas; these photographs were taken 7 minutes after those of Fig. 6. Here the narrowest bandwidth is about 3 mc [Fig. 7(n)]. This is narrower than the predicted value of 5 mc. Fig. 8 shows the received signal on an 8′ antenna a few minutes later. Due to the lower gain of the 8′ antenna, the signal level is closer to the noise level of the system. However, one notes the narrowest bandwidth to be about 3 mc [Fig. 8(n)].

The photographs show that the medium introduces selective fading which bears a close similarity to the fading which occurs on a line-of-sight path during anomalous propagation conditions.[4] It raises the question as to how many layer-like discontinuities are present at one time. The hypothesis that there are a limited number of regions of discontinuities does not necessarily contradict the fact that the received signal is Rayleigh-distributed as a function of time since in each region a few layers could change both in amplitude and position with time, therefore giving a Rayleigh distribution.

To get a more quantitative answer as to the bandwidth of the medium and its relation to the antennas used, an analysis is being made to determine the cross-correlation function with frequency as well as the distribution of bandwidth with time.[5] A more realistic definition of bandwidth in terms of change of amplitude as a function of frequency change will be used. Nevertheless, the prediction of a bandwidth of the order of 4–7 mc is consistent with the experimental data.

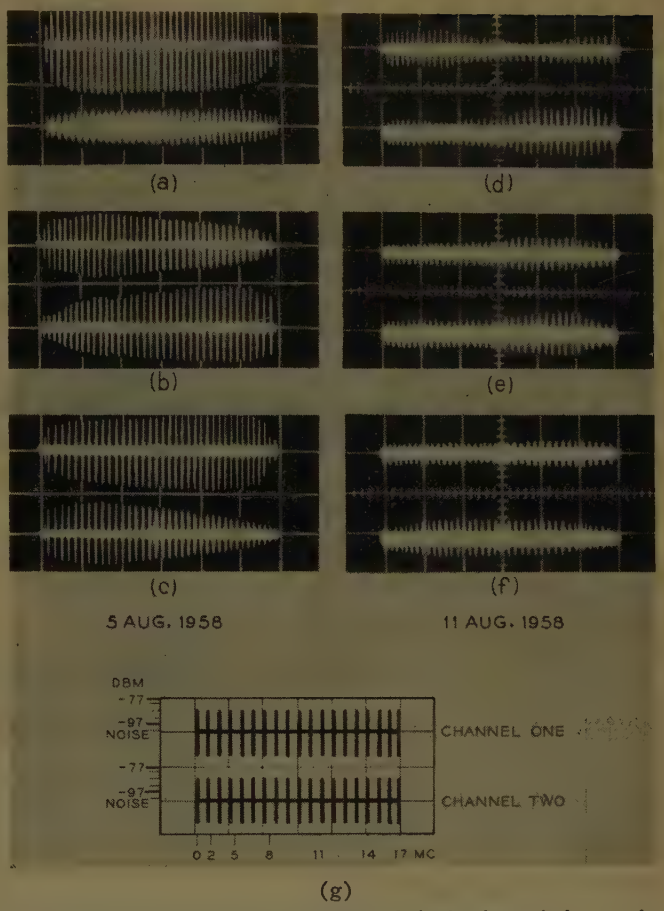


Fig. 9—Sweep-frequency photographs of single channels in a twin-feed diversity system with horizontally disposed feeds—4110 mc. (g) calibration.

Results—Twin-Feed Diversity Experiments

Twin-feed diversity experiments, using cw transmissions, have shown that signals from the two received channels are uncorrelated most of the time at 4 kmc.[6] To examine the frequency correlation between two diversity channels, tests were made using the 28′ transmitting antenna with a single feed and the 28′ receiving antenna with two horizontally disposed feeds. Simultaneous recordings were made on the two channels using a dual trace oscilloscope. Photographs were taken with a Polaroid camera with 1/25 second exposure. Fig. 9 shows several selected frames which illustrate the amplitude-frequency characteristics; Fig. 9(g) shows the calibration for the photographs. Fig. 9(a)–(c) were recorded on a day when the bandwidth was rather broad, whereas the remaining were taken on a more representative day. In Fig. 9(a) and (e), the amplitude-frequency curves are well correlated; on the remaining ones there is less correlation. In all the runs the signals between the two channels at any particular frequency were uncorrelated. An examination of several hundred frames shows that a frequency effect does occur between the two diversity channels when the frequency separation is of the order of several megacycles at 4 kmc.

If the received channels are combined in a switch-type combiner, the percentage of time that the switched signal (at one frequency) spends below a certain level will decrease and the bandwidth will be broader.

Conclusions

Selected sequences of sweep-frequency photographs taken on a 171 mile beyond-the-horizon circuit have been shown; these exhibit selective fading characteristics typical of multipath propagation. The results of a calculation based on the geometry of the common volume formed by the transmitting and receiving antennas predict a bandwidth of from four to seven megacycles. The bandwidth thus calculated is consistent with the one obtained from the experimental data.

A look at the simultaneous recordings of two channels from a twin-feed diversity system shows that the amplitude-frequency functions of the two channels are uncorrelated in most instances.

Acknowledgment

The author would like to acknowledge the assistance of Ralph A. Semplak, who built and maintained a large part of the equipment and took much of the data.

[4] A. B. Crawford and W. C. Jakes, Jr., "Selective fading of microwaves," *Bell Sys. Tech. J.*, vol. 31, pp. 68–90; January, 1952.

[5] A. B. Crawford, D. C. Hogg, and W. H. Kummer, "Studies in tropospheric propagation beyond the horizon," *Bell Sys. Tech. J.*, vol. 38, pp. 1067–1178; September 1959. (See Sec. VIII.)

[6] W. H. Kummer, "Twin-feed Diversity Studies in Beyond-the-Horizon Propagation," presented at URSI Meeting, Washington, D. C.; April 26, 1958.

communications

Geometrical Optics Approximation of Near-Field Back Scattering*

F. SHEPPARD HOLT†

IN the radar equation

$$P_r = \left(\frac{P_t G}{4\pi R^2}\right)\left(\frac{\sigma}{4\pi R^2}\right)\left(\frac{G\lambda^2}{4\pi}\right) \tag{1}$$

where

P_r = received power
P_t = transmitted power
G = antenna gain
R = range
σ = back-scattering cross section of the target
λ = wavelength

it is assumed that the path attenuation from the radar to the target is proportional to $1/R^2$. This is a valid assumption provided the target is in the far-field of the radar. The back-scattering cross section σ is defined by

$$\sigma = \frac{4\pi R^2 S_r}{S_i} \tag{2}$$

where S_i is the power density incident at the target and S_r is the reflected power density measured at the radar. If the return path attenuation is proportional to $1/R^2$, that is, if the radar is in the far-field of the target, and if the environmental conditions, the wavelength, and the target aspect are fixed, then σ as given by (2) will be essentially constant. However, if the radar is not in the far-field of the target then (2) will in general require σ to be a function of the range R.

* Manuscript received by the PGAP, April 29, 1959; revised manuscript received, June 15, 1959.
† AF Cambridge Res. Center, Bedford, Mass.

As an example, consider the reflection at normal incidence from a perfectly conducting infinite plane in the far-field of a radar. The radar will always be in the near-field of the infinite plane and thus it is expected that σ will be a function of R. The power density ratio S_r/S_i, as determined by geometrical optics, is equal to 4 for the infinite plane and hence by (2)

$$\sigma = \pi R^2, \tag{3}$$

a function of R as expected. Substitution of (3) into (1) leads to the result

$$P_r = \frac{P_t G^2 \lambda^2}{64\pi^2 R^2}$$

which can be verified directly by the method of images.

Consider the reflection from a convex, doubly curved, perfectly conducting surface in the far-field of a radar. Assume that the surface is illuminated over many Fresnel zones and that at the specular reflection point the principal radii curvature ρ_1 and ρ_2 are large compared to λ. In this case geometrical optics predicts a back-scattering cross section as follows:

$$\sigma = \frac{\pi \rho_1 \rho_2 R^2}{(R+\rho_1)(R+\rho_2)}, \tag{4}$$

where R is the range from the radar to the specular reflection point on the doubly curved surface. To check the validity of this formula consider the two limiting cases: 1) $\rho_1 = \rho_2 \to \infty$ and, 2) $\rho_1 = \rho_2 = \rho$, $R \to \infty$. Case 1 leads to the back-scattering cross section of a perfectly

conducting infinite plane, and the formal result of applying the indicated limiting operations to (4) is $\sigma = \pi R^2$ or precisely (3). Case 2 leads to the determination of σ for a perfectly conducting sphere whose radius ρ is large compared to the wavelength λ, but very small compared to the range R. The indicated limiting operations in this case when applied to (4) lead to the well-known result

$$\sigma = \pi\rho^2.$$

It is interesting to apply (4) to the case of the moon illuminated by a radar on the earth assuming that the moon is a smooth sphere with uniform reflectivity coefficient. At a range of 240,000 miles and with a radius of 1090 miles any point on the surface of the earth is in the near-zone of the moon at all frequencies above 4.9 kc. This assumes the minimum far-field range to be $2D^2/\lambda$ where D is the diameter of the moon. Application of (4) leads to the result

$$\sigma = 0.991\sigma_0,$$

where σ_0 is the *far-field* back-scattering cross section of a smooth spherical moon of uniform reflectivity coefficient. Therefore it is evident *under the assumptions made* that the fact that the earth is not in the far-field of the moon has very little effect (less than 1 per cent) upon the back-scattering cross section of the moon.

Scanning Antenna Arrays of Discrete Elements*

E. A. BLASI† AND R. S. ELLIOTT‡

RADAR systems designers currently are showing considerable interest in scanning antennas composed of discrete elements, such as dipoles, slots, helixes, etc. The principal attraction appears to be more positive control over the aperture distribution than is possible with reflectors, lenses, and other continuous aperture types. Depending on the method of feeding the discrete elements, this advantage may prove entirely illusory. It is the purpose of this note to point up some of the difficulties which are encountered in a design approach using discrete elements.

When linear polarization is desired, dipoles and slots are frequent choices as the radiating element. For a planar array, the two are equivalent via Babinet's principle, and for this reason we shall confine the discussion to dipoles. Using Ronald King's data for self- and mutual impedances of parallel identical dipoles[1] in addition to the usual assumptions about superposition and images, it is a tedious but straightforward process to compute the driving point impedance of a single dipole in an array.[2] For example, the case of a 5×3 array of dipoles has been considered. The arrangement is shown in Fig. 1. The dipoles are all $\lambda_0/2$ on centers and $\lambda_0/4$ in front of a ground plane, with λ_0 the mid-band wavelength. It is assumed that the amplitude of excitation is uniform and that the phase of excitation is uniform progressive. Therefore, this array approximates quite closely a 5×3 internal segment of a large array, and the driving point impedance of the central element is close to the driving point impedance values for a majority of the dipoles in a large array. An $R-X$ plot of the driving point impedance of the central element is displayed in Fig. 2. The two parameters are frequency (via the wave-

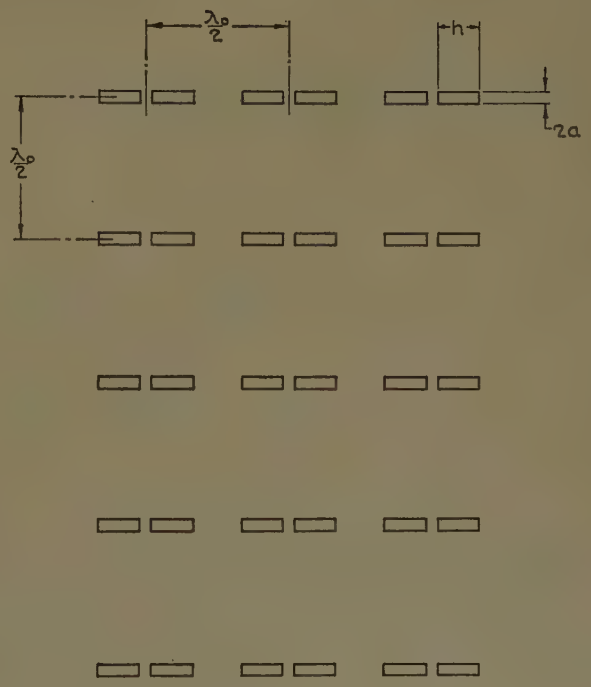


Fig. 1—Geometry of 5×3 dipole array.

* Manuscript received by the PGAP, May 25, 1959.
† Lockheed Missiles and Space Div., Sunnyvale, Calif.
‡ Dept. of Engrg., University of California, Los Angeles, Calif.
[1] R. W. P. King, "The Theory of Linear Antennas," Harvard University Press, Cambridge, Mass.; 1956.
[2] E. A. Blasi, "Effects of Mutual Interactions on the Design of Various Dipole Arrays," Hughes Res. Labs., Culver City, Calif., Tech. Memo. No. 336; December, 1953.

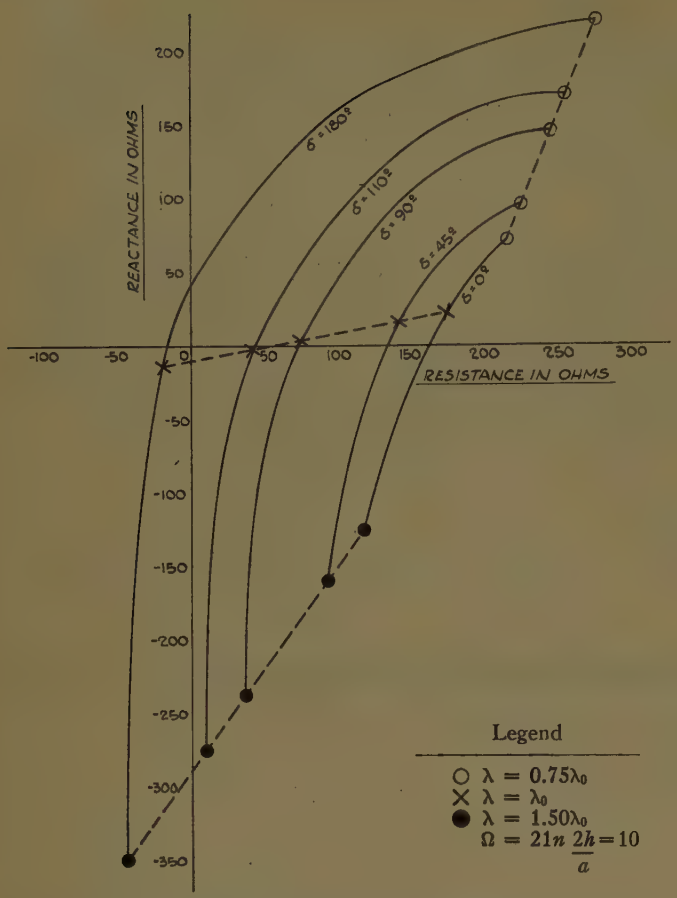


Fig. 2—Driving point impedance of a central dipole element vs frequency and beam position.

length λ) and interelement phaseshift, δ. At the midband wavelength λ_0, the beam scans in the E-plane from broadside ($\delta = 0°$) to endfire ($\delta = 180°$). Using impedance stability as a criterion, the array dipole is seen to be not only narrow-band, but also narrow-range when scanning is considered. Similar results are obtained for scanning in the H-plane, and thus for two-dimensional beam motion.

For very large arrays, with all elements individually fed (or terminated), the principal effect of these impedance variations is on the gain of the antenna, with good maintenance of pattern shape. When conventional standing-wave and traveling-wave transmission lines are employed, the effect on pattern is disastrous and elaborate correction circuits must be added to the system. In either case, it is interesting to observe from Fig. 2 that there exists a "cut-off" point at the lower frequency when approaching the endfire condition, since the driving point impedance tends towards a negative value. It has been the experience of several experimental investigators to actually measure these negative values.

Experimental results confirm the calculations just presented. Similar calculations for other element geometries, such as the helix, are extremely difficult and have not been attempted. However, preliminary experimental results on helixes are equally sobering, and our suspicion is growing that, for many applications, a handsome price is to be paid for the aperture control gained by going to discrete elements.

On the Use of Uniform Circular Arrays to Obtain Omnidirectional Patterns*

TA-SHING CHU†

MANY omnidirectional antennas have been designed in the form of circular arrays of equally-spaced identical elements fed in phase. The term "omnidirectional" is restricted here to azimuthal patterns, *i.e.*, patterns in the plane of the array. It is the purpose of this study to determine the relationship between the number of radiating elements and the fluctuations in the azimuthal pattern as the radius of the circular array varies. Of particular interest are the conditions under which a good omnidirectional pattern can be obtained from such an array. If there are S identical elements fed in phase so that a rotation of the array through $2\pi/S$ radians about an axis perpendicular to its plane does not alter the appearance of the array, it is evident that each point in the radiation pattern must be repeated $(S-1)$ times. As the number of sources increases, decreasing the angular spacing, it is to be expected that the pattern variation between corresponding points will generally decrease. The influence of the pattern of a single element on the array pattern fluctuation is also of interest.

The problem of obtaining good omnidirectional patterns from circular arrays has been touched upon in a number of earlier papers. A qualitative discussion of the omnidirectional properties of circular arrays has been

* Manuscript received by the PGAP, April 17, 1959. This work was sponsored in part by the U. S. Army, Engrg. Lab., Fort Monmouth, N. J.

† Antenna Lab., Dept. of Elec. Engrg., The Ohio State University, Columbus, Ohio.

given by Hansen and Woodyard[1] and Hansen and Hollingsworth.[2] The radiation from dipoles mounted on a conducting cylinder has been analyzed by Carter[3] for cases likely to be of interest in the design of radio and television antennas. A synthesis method for circular arrays has been described by Taylor.[4] The radiation from circular arrays of dipoles has been treated by Page[5] and Knudsen.[6] In this paper, the method of analysis given in Knudsen's paper has been extended to circular arrays of arbitrary, identical elements; the detailed discussion, however, is restricted to elements having symmetrical patterns. The pattern fluctuations are shown in graphical form for several types of elementary patterns. Since these results were to be used in the microwave region, curves are given for larger values of S than presented in the earlier papers.

The problem to be considered is the radiation from S elements arranged in a circle. The elements are identical, equally-spaced, and fed in phase. Since their far-zone patterns are of interest, the location of the elements may be described simply by ideal point sources. A spherical coordinate system (r, θ, ϕ) is centered at the origin of the ring array shown in Fig. 1. The plane of the array is perpendicular to the polar axis so that the coordinates of the jth source are $(a, \pi/2, u_j)$, where $u_j = j(2\pi/S)$ with $j = 1, 2, 3, \cdots, S$.

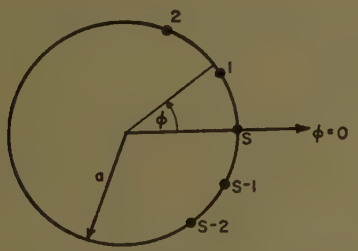


Fig 1—A uniform circular array.

The symmetrical pattern $F(\phi')$ of the jth point source of the array can be represented by the Fourier cosine series

$$F(\phi') = \sum_{n=0}^{\infty} a_n \cos n\phi' = \sum_{n=0}^{\infty} A_n \cos^n \phi' \qquad (1)$$

where $\phi' = 0$ at $\phi = u_j$ and the constants A_n, which may be complex, are functions of θ. The condition under which $F(\phi')$ is determined is important if the interaction between elements is to be taken into account properly and the principle of superposition accurately applied.[7]

However, for most design work, approximations can be made; usually, the pattern of a single element is simply taken to be the array pattern for $S = 1$. Applying the superposition principle, the total far-field pattern for S elements is

$$\Phi = \sum_{j=1}^{S} \left[\sum_{n=0}^{\infty} A_n \cos^n (\phi - u_j) \right] e^{iZ \cos (\phi - u_j)}$$

$$= \sum_{n=0}^{\infty} A_n (-i)^n \frac{d^n}{dZ^n} \left[\sum_{j=1}^{S} e^{iZ \cos (\phi - u_j)} \right] \qquad (2)$$

where $Z = ka \sin \theta$ and for $n = 0$ there is no differentiation with respect to Z. After some manipulation, in which the formulas

$$e^{iZ \cos (\phi - u_j)} = \sum_{m=0}^{\infty} (2 - \delta_{0m}) i^m J_m(Z) \cos m(\phi - u_j) \qquad (3)$$

and

$$\sum_{j=1}^{S} e^{imu_j} = \sum_{j=1}^{S} e^{i(2\pi mj/s)}$$

$$= \begin{cases} S & \text{for } \dfrac{m}{S} = p \\ & \text{where } p = 0, \pm 1, \pm 2, \cdots \\ 0 & \text{in any other case} \end{cases} \qquad (4)$$

are used,

$$\Phi = S \sum_{p=0}^{\infty} \left[\sum_{n=0}^{N} A_n i^{sp-n} \left(\frac{d^n}{dZ^n} J_{sp}(Z) \right) \right] \cdot (2 - \delta_{0(sp)}) \cos sp\phi. \qquad (5)$$

In the above equation, the infinite sum on n has been replaced by a finite sum, since any practical Fourier representation of a pattern function would employ a finite number of N terms. The summation with respect to p is rapidly convergent for a large number of elements in an array of small radius, because sp becomes larger than Z within a few terms and the order of the Bessel function exceeds its argument. If the number of terms N in the Fourier representation is less than the number of elements S, then Φ may be approximated by

$$\Phi \approx S \sum_{n=0}^{N} A_n (-i)^n \frac{d^n}{dZ^n} [J_0(Z) + 2(i)^s J_s(Z) \cos s\phi]. \qquad (6)$$

The approximation is most accurate for patterns with small fluctuations.

[1] W. W. Hansen, and J. R. Woodyard, "A new principle in directional antenna design," Proc. IRE, vol. 26, pp. 333–345; March, 1938.

[2] W. W. Hansen, and L. M. Hollingsworth, "Design of flat-shooting antenna arrays," Proc. IRE, vol. 27, pp. 137–143; February, 1939.

[3] P. S. Carter, "Antenna arrays around cylinders," Proc. IRE, vol. 31, pp. 671–697; December, 1943.

[4] T. T. Taylor, "A synthesis method for circular and cylindrical antennas composed of discrete elements," IRE Trans. on Antennas and Propagation, vol. AP-3, pp. 251–261; August, 1952.

[5] H. Page, "Ring-aerial systems—minimum number of radiators required," *Wireless Engr.*, vol. 25, pp. 308–315; 1948.

[6] H. L. Knudsen, "Radiation from ring quasi-arrays," IRE Trans. on Antennas and Propagation, vol. AP-4, pp. 452–472; July, 1956.

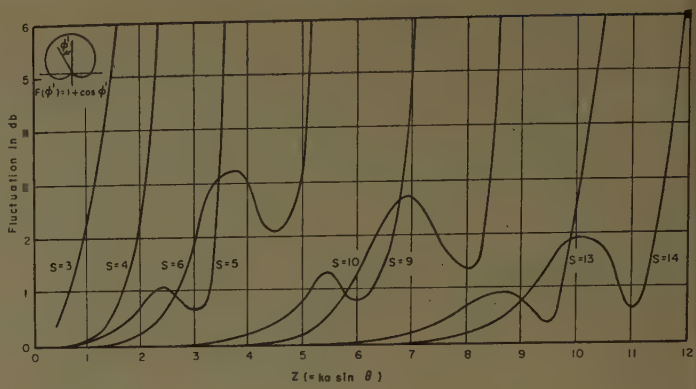
[7] J. N. Hines, *et al.*, "On the design of arrays," Proc. IRE, vol. 42, pp. 1262–1267; August, 1954.

In order to examine some of the properties of uniform circular arrays, the previous expression is applied to several simple cases, and the fluctuation in the azimuthal pattern, defined as the ratio of maximum $|\Phi|$ to minimum $|\Phi|$, is determined. In Fig. 2(a), curves for different S have been plotted as a function of $ka \sin \theta$ for an array composed of sources having a pattern of the form $(1+\cos \phi')$, and in Fig. 2(b), similar curves have been plotted for an array composed of sources having a pattern of the form $(1+\frac{1}{2} \cos \phi')$. They are typical examples of arrays composed of broadly directional elements and qualitatively indicate the general behavior of such arrays. For fluctuations of less than 1 db, the two sets of curves compare closely. However, as would be expected, the larger fluctuations are more sensitive to the element pattern; *e.g.*, a 5-db fluctuation with S fixed results in differences in the array radius of up to 20 per cent. On the other hand, the results for sources with an omnidirectional pattern are plotted in Fig. 2(c) and are seen to be considerably different from the previous curves. At certain values of $ka \sin \theta$, the fluctuation becomes very large independently of S as indicated by the dotted lines in Fig. 2(c). These occur for $J_0(Z)=0$, as may be seen from (6), noting that in this case, $N=0$. It is clear that arrays of broadly directional elements do not have this undesirable property, so that a satisfactory pattern can be obtained over a desired band of frequencies by choosing the proper number of sources. This, of course, assumes that the pattern of the source does not deteriorate in the frequency band. In the case of the omnidirectional elements, an odd number of sources gives less azimuthal pattern fluctuation than an even number. This is not true for arrays composed of broadly directional elements.

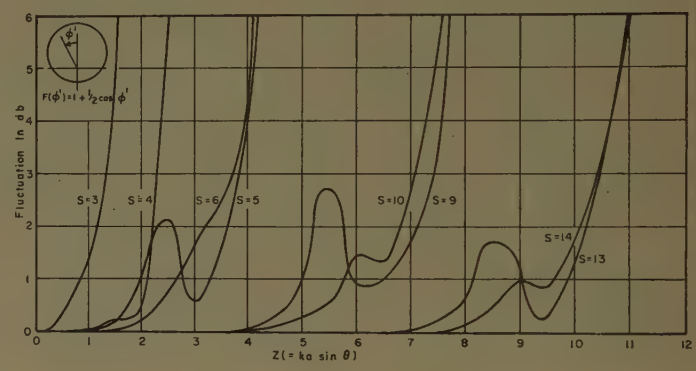
Eq. (6) indicates that as the number of elements is increased for a given Z, in general, the fluctuation in the azimuthal pattern decreases. Moreover, for the array radius less than the value corresponding to the first maximum in the curve, there is less fluctuation at the higher elevation angles, because the fluctuation is a function of $ka \sin \theta$.

In Figs. 2(a) and 2(b), it is interesting to observe that good omnidirectional patterns can be achieved for arrays consisting of more than a few elements, if the elements are spaced about one half-wavelength apart or less. This agrees well with an intuitive extension of the half-wavelength criteria for the existence of a single, well-defined maximum in the radiation pattern of a linear array,[8] and with another prediction of a qualitative nature which can be deduced from the work of Taylor.[4]

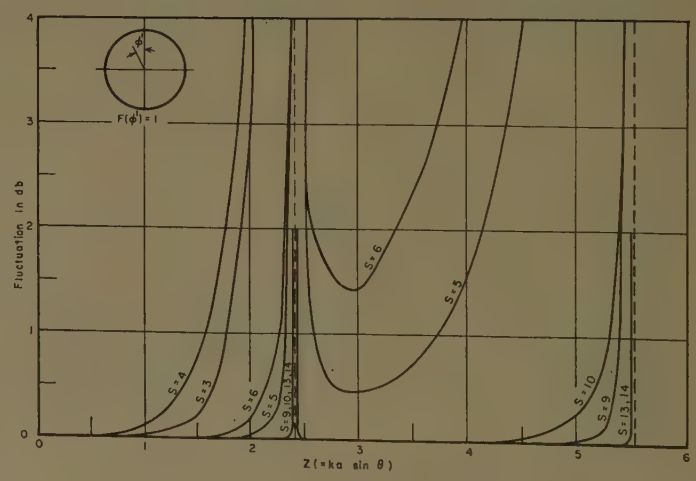
[8] This was pointed out by C. Eason of the U. S. Army Signal Research and Development Laboratory in an informal conversation.



(a)



(b)



F(φ') Is The Pattern Of A Single Element
s = The Number Of Elements

$$\text{Fluctuation} = 20 \log_{10} \left[|\Phi|_{\max.} / |\Phi|_{\min.} \right]$$

(c)

Fig. 2—The fluctuation in the azimuthal pattern of a uniform circular array.

The effects of progressive phase shift between identical elements has been treated by Knudsen[6] for the case of arrays composed of omnidirectional elements. His analysis can easily be generalized using the method described here.

Status of Tropospheric Extended Range Transmission*

K. BULLINGTON†

FREQUENTLY reporters, and sometimes even engineers, ask such questions as: who invented tropospheric extended range transmission, and when? Or, what single experiment led to its discovery? There is no easy answer to any of these questions. It is simply a matter of how far back in history we want to go and what constitutes a beginning.

In the case of extended-range or beyond-horizon tropospheric transmission, Marconi on at least one occasion received 400-mc signals over a distance of about 150 miles. That was more than twenty-five years ago and there may even have been someone else who predated him. Marconi had a hunch or faith that UHF could be useful beyond the horizon, but he did not have either the data on which to build a good theory or the facilities with which to get the data.

In the last twenty-five years, many people have contributed to our present understanding of beyond-horizon transmission. The use of higher power at higher frequencies has provided an increasing amount of data which showed that the old ideas were incomplete and provided both the incentive and the material for new theories. Some of the measured programs were by-products of other work, some were definitely planned to check a theory and some were designed primarily to gather statistical data for engineering purposes. During these years, some data have had to be reinterpreted and some theories have had to be modified, but this is a necessary part in understanding any phenomenon. Most of the early work was summarized or at least mentioned in the October, 1955 issue of the PROCEEDINGS OF THE IRE.

There now seems to be reasonable agreement on the major characteristics of tropospheric transmission, which has been called by various names such as extended range, transhorizon, beyond horizon, scatter or that unwieldy FPTS.

The principal characteristic is that the received power decreases rapidly with distance, perhaps as the 7th or 8th power of distance. In spite of this attenuation, useful signals have been measured out to 600 miles or more.

The average signal level is lower in the Arctic than in the tropics, but useful signals can be received even in the coldest and driest climates. In the middle latitudes, there is a seasonal variation of ± 10 db or more with the highest signals in the summer. The variation with season and latitude depends on atmospheric refraction and is more or less proportional to our old friend, k, the effective earth's radius factor.

The short-term fading is essentially random; that is, it follows a Rayleigh distribution, and some form of space or frequency diversity is necessary. Antennas as large as 40 wavelengths in diameter can achieve essentially their full gain. Larger antennas fail to realize their full plane-wave gain because of a lack of phase coherence. However, useful increases in gain may be realized up to about 100 wavelengths, and these larger antennas may also be helpful in increasing the useful bandwidth. While some lack of correlation does exist at a separation of 40 wavelengths, full diversity advantage requires a separation of 100–150 wavelengths.

Although the fading is rapid, the significant components of the signal in practical systems are separated in time by only fractions of a microsecond. As a result, the available bandwidth is sufficient to provide an acceptable grade of black and white television service for at least one link. Alternatively, 100 or more telephone channels can be provided on a single radio carrier. High-quality voice transmission and even data transmission can be obtained with high reliability, providing the design has adequate signal-to-noise margin.

Any attempt to obtain maximum distance, maximum bandwidth and maximum quality or reliability all in one design is doomed to failure. Any one of these desirable features can be maximized, but at the expense of at least one of the other features.

The received power decreases somewhat as the frequency increases, but frequencies from less than 100 mc to more than 5000 mc may be useful. This is illustrated in Fig. 1, which shows the estimated range of a single link circuit for various values of antenna size and frequency. For example, with 120-foot antennas, the optimum frequency is below 200 mc, while with 30-foot antennas, the optimum frequency is below 1000 mc. The optimum frequency will be increased when lower noise figures can be maintained at the higher frequencies. The chart assumes that the transmitter power, bandwidth and reliability are held constant. These estimates also assume the use of dual diversity reception; the use of quadruple diversity increases the estimated ranges by about 15 per cent.

Sometimes the questions are asked: how does the energy get beyond the horizon? What is the real mechanism? The answer is not yet entirely clear. Equations have been evolved from both experimental data and theory, and gradually these two have come into reasonable agreement. The physicist is usually satisfied with the word "scatter," but the layman interprets scattering

* Manuscript received by the PGAP, November 12, 1958; revised manuscript received, April 24, 1959. This paper is the substance of a talk at the National Symposium on Extended Range Transmission in Washington, D. C., October 7, 1958. Its purpose is to summarize the present status of extended range tropospheric transmission with brief mention of the history, principal characteristics, nature of the mechanism and general usefulness.

† Bell Telephone Laboratories, Inc., Murray Hill, N. J.

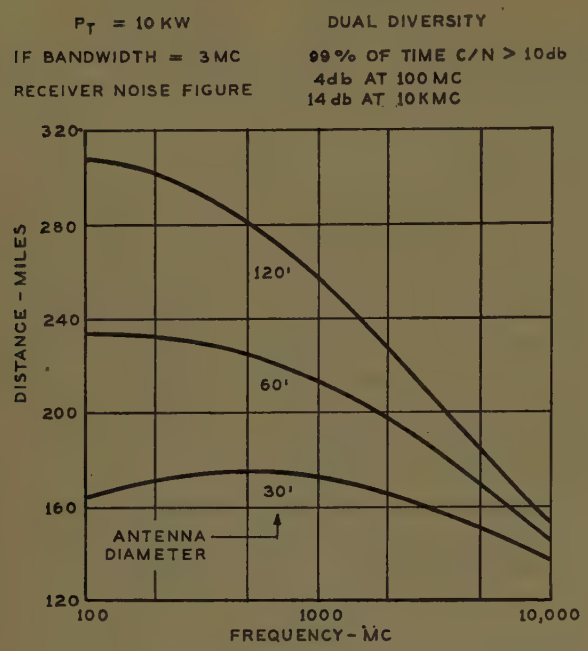


Fig. 1—Estimated range for beyond-horizon transmission.

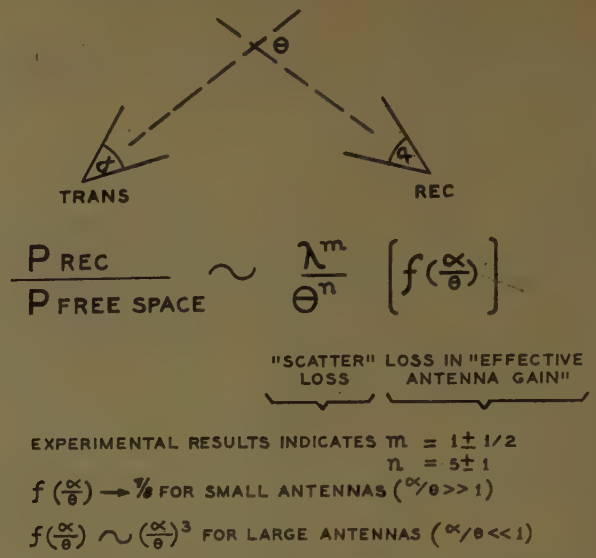


Fig. 2—Beyond-horizon tropospheric transmission.

as spreading in more or less all directions. The experimental characteristics fit more closely to the layman's concept of reflection rather than scattering but here also there may be some conceptual difficulties.

A simplified equation for the loss in power relative to that expected in free space is shown in Fig. 2. The experimental results show a small wavelength dependence of about $m=1$. This is difficult for the scattering theory to explain unless the blobs are squashed into layers. The distance dependence of $n=5\pm1$ is more than can be accounted for by simple scattering alone. A further assumption is needed regarding the way in which the average variations in dielectric constant change with height above the ground. Almost any value for n can be obtained by assuming the right change in dielectric constant with elevation; and this important parameter is difficult to measure and perhaps variable with time. Moreover, the antenna factor $f(a/\theta)$ can confuse the issue by providing an additional exponent of 1, 2, or 3 when very narrow-beam antennas are used.

For engineering purposes, it is sufficient to emphasize the equation that fits the experimental results and to minimize the fine shades of meaning in the interpretation of the equation. While this procedure may not completely satisfy our desire for basic knowledge, it does have some good historical precedents. For example, Isaac Newton discovered a formula that accounted for the motion of the planets, but he did not explain the mechanism of gravitation. Some of his contemporaries saw nothing new in this approach because it showed only the how and not the why of gravitation. Gradually, most people stopped asking why gravitation occurred and simply accepted the formula. As radio engineers, we may ultimately need to take a similar position in regard to tropospheric transmission.

Another question frequently asked is: what is beyond-horizon tropospheric transmission good for and where will it be used? As already mentioned, it is capable of relatively wide-band transmission. It is being used primarily in the Arctic and over water links, and its use will probably increase in areas of difficult terrain. On the other hand, in well-populated regions with a good network of roads and power lines, beyond-horizon transmission cannot compete economically with comparable line-of-sight microwave systems. In addition, the high power involved may cause interference into other services unless special precautions are taken.

Consequently, beyond-horizon transmission is complementary to and is not a substitute for line-of-sight systems. It is very useful in areas where line-of-sight transmission is difficult, but the circuits are relatively expensive and, hence, cannot compete in regions that are more favorable to line-of-sight systems.

Contributors

Carlos M. Angulo (S'50–A'52–M'52–SM'56), for a photograph and biography please see page 304 of the July, 1959, issue of these TRANSACTIONS.

❖

D. K. Bailey, for a photograph and biography please see page 387 of the October, 1958, issue of these TRANSACTIONS.

❖

Doran J. Baker was born in Salt Lake City, Utah, on October 11, 1931. He received the B.S. degree in 1953, and the Ph.D. degree in electrical engineering in 1956, both from the University of Utah, Salt Lake City.

D. J. BAKER

From 1952 to 1956 he was a research associate with the Upper Air Research Laboratories, University of Utah, engaged in the study of the electromagnetic properties of the ionosphere by the use of high-altitude sounding rockets. As a commissioned officer in the U. S. Air Force, he served from 1956 to 1958 as an electronics engineer with the Ionospheric Physics Laboratory of the Geophysics Research Directorate, Air Force Cambridge Research Center, Bedford, Mass. From 1958 to 1959 he was employed as a physicist by the Thermal Radiation Laboratory of the same Directorate, where he was involved in the study of nuclear thermal radiation. Since 1959 he has been an assistant professor at the Utah State University, Logan, where he is assistant director of the Electro-Dynamics Laboratory.

Dr. Baker is a member of Tau Beta Pi, Sigma Xi, Sigma Pi Sigma, Phi Kappa Phi, the Air Force Association, and the American Meteorological Association.

❖

Richard H. T. Bates was born in Sheffield, England, on July 8, 1929. He received the B.Sc. (Eng.) degree from London University in 1952.

R. H. T. BATES

He joined the Guided Weapons Department of Vickers-Armstrong (Aircraft) Ltd. of Weybridge, England, in 1952, and in 1955 he transferred to the Aerial Section of Decca Radar Co., Ltd., of Tolworth, Surrey, England. From there he moved to Canada in 1957 to join the Electronics Division of Canadian Westinghouse Co. Ltd., Hamilton, Ontario, where he is presently employed.

Mr. Bates ia a graduate member of the Institution of Electrical Engineers.

❖

William S. C. Chang (S'56–M'58), for a photograph and biography please see page 304 of the July, 1959, issue of these TRANSACTIONS.

❖

Orval R. Cruzan was born near Cushing, Okla., on July 4, 1914. He received the B.S. degree in physical science in 1937 from Oklahoma State University, Stillwater. In 1941 he joined the staff of the National Bureau of Standards as a physicist in the field test section of the Ordnance Development Division. In the same organization, he worked as a mechanical engineer from 1945 to 1948. From 1948 to 1952 he worked again as a physicist, and as a mathematical physicist from 1952 to 1954. In 1953, he became a member of the staff of the Diamond Ordnance Fuze Laboratories, Washington, D. C., through the transfer of the Ordnance Development Division to the Ordnance Corps, Department of the Army. From 1954 to the present, he has worked as an applied mathematician.

O. R. CRUZAN

Mr. Cruzan is a member of the American Physical Society.

❖

Lorne H. Doherty (S'42–A'46–M'53), for a photograph and biography please see page 204 of the April, 1959, issue of these TRANSACTIONS.

❖

John D. Dyson (S'49–A'52–M'58) was born on August 9, 1918, in Lemmon, S. D. He received the B.S. degree in economics and the B.S. degree in electrical engineering in 1940 and 1949, respectively, from South Dakota State College, Brookings. In 1950, he received the M.S. degree, and in 1957 the Ph.D. degree, both in electrical engineering from the University of Illinois, Urbana.

J. D. DYSON

From 1940 to 1941 he was employed as a statistician; from 1941 to 1946 he served on active duty with the U. S. Army. He was a part-time instructor at South Dakota State in 1949 and was on the research staff of the Sandia Corporation, Albuquerque, N. M., from 1951 to 1952. Since October, 1952, he has been on the staff of the antenna laboratory of the University of Illinois, where he is now a research assistant professor.

Dr. Dyson is a member of Sigma Xi, Eta Kappa Nu, Sigma Tau, and Pi Mu Epsilon.

❖

Hermann W. Ehrenspeck was born in Kusel, Germany, on February 27, 1912. He received the B.S. and M.S. degrees in engineering from the Institute of Technology of Munich, Germany, in 1935 and 1937, respectively, and the Ph.D. degree from the same institute in 1947.

H. W. EHRENSPECK

He was employed in the Air Research Institute of Oberpfaffenhofen, Germany, from 1937 to 1945, where he worked on directional antennas, on propagation of electromagnetic waves, and on antennas for direction finding. From 1946 to 1953 he was in charge of the research laboratory of an electrical components company. Since 1953 he has been employed as a research physicist in the Electromagnetic Radiation Laboratory of the Air Force Cambridge Research Center, Air Research and Development Command, Bedford, Mass.

❖

K. Fränz was born in Berlin, Germany, on February 12, 1912. He received the Ph.D. degree in philosophy from Berlin University, Germany, in 1936, and the Dr.ing. habil degree from Technical High School Berlin-Charlottenburg in 1943.

K. FRÄNZ

He joined the optical laboratory of Osram, Berlin in 1931, the Heinrich-Hertz-Institute, Berlin-Charlottenburg, in 1933, the Telefunken GmbH in 1946, Fabricas Militares of Buenos Aires in 1948, and the Argentine Atomic Energy Commission in 1954. From 1949 to 1956 he was a pro-

fessor at Buenos Aires University, Argentina. At present he lectures at Tübingen University, Germany, and works at the Telefunken Research Institute in Ulm, Germany. He has published papers, or books, on propagation, antennas, linear and nonlinear noise, high-frequency heating microwave circuits, feedback and control, reactor simulators, and nuclear instrumentation.

Dr. Fränz is a member of the Nachrichtentechnische Gesellschaft of Germany and the Physical Society of London.

❖

Leonard O. Goldstone (M'51–SM'55) was born in Ogdensburg, N. Y., on February 15, 1922. He received the B.E.E. degree from Clarkson College, Potsdam, N. Y., in 1947, and the M.E.E. and D.E.E. degrees from the Polytechnic Institute of Brooklyn, N. Y., in 1950 and 1957, respectively. From 1947 to 1954 he was an instructor and then assistant professor in the Electrical Engineering Department of the Polytechnic Institute of Brooklyn, Brooklyn, N. Y.

L. O. Goldstone

From 1954 to 1957 he was a research associate at the Microwave Research Institute, where he was engaged in research on propagation in unconventional waveguides with applications to traveling-wave antennas. In 1957, he returned to the faculty of the Polytechnic Institute of Brooklyn, where he is now an associate professor of electrical engineering. He is also presently serving as a research consultant to Kahn Research Laboratories, Inc.

Dr. Goldstone is a member of Eta Kappa Nu and Sigma Xi.

❖

Obed C. Haycock attended Utah State University, Logan, and the University of Utah, Salt Lake City, receiving the B.S.E.E. degree in 1925. He received the M.S.E.E. degree in 1931 from Purdue University, Lafayette, Ind., and continued graduate study at the University of Michigan, Ann Arbor, and the University of Utah.

O. C. Haycock

From 1925 to 1926 he was employed by Westinghouse Manufacturing Co., East Pittsburgh, Pa., working on automatic substation control equipment. In 1925 he joined the Department of Electrical Engineering, University of Utah, where he has been a full professor since 1947. From 1947 to 1957 he was an associate director, and since 1957 a director, of the Upper Air Research Laboratories at the University of Utah.

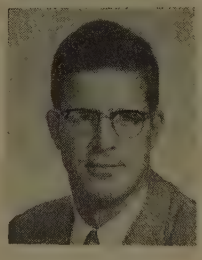
During the summers of 1923 to 1929, he was employed by the U. S. Bureau of Reclamation, Wilbert Mining Co., and Jackson Power Co., working on power house operation, design and installation of power transmission lines and installation of power house equipment. In 1944 to 1945 he was given a leave of absence from the University of Utah to Rutgers University, New Brunswick, N. J., in order to participate as research engineer in research on jungle acoustics being carried out in the Panama Canal zone.

His early research was concerned with the measurement of the attenuation of electromagnetic waves through the earth and engineering development of antennas. Since 1947 his research has been with the Upper Air Research Laboratories, engaged in measuring electron densities and gas densities in the upper atmosphere. In conjunction with the Air Force Cambridge Research Center, he has participated in five V2, twenty-three Aerobee and three Nike-Cajun rocket experiments.

Mr. Haycock is a member of the AIEE, Tau Beta Pi, Sigma Xi, and the ASEE.

❖

Richard C. Honey (S'48–A'53–SM'57) was born on March 9, 1924 in Portland, Ore. He received the B.S. degree in physics from the California Institute of Technology, Pasadena, in 1945. After serving as a radio technician in the U. S. Navy, he attended Stanford University, Stanford, Calif., where he received the E.E. degree in 1950, and the Ph.D. degree in electrical engineering in 1953. He was a member of the microwave oscillator project at Stanford Electronics Research Laboratory from 1948 to 1952.



R. C. Honey

In 1952, he joined the staff of Stanford Research Institute, Menlo Park, Calif., where he has worked on microwave antenna systems and components and on parametric amplifiers.

Dr. Honey is a member of the Scientific Research Society of America and Sigma Xi.

❖

Ming-Kuei Hu (S'48–A'51–SM'56) was born on May 25, 1918, in Anhwei, China. He received the B.E.E. from National Central University, Chungking, China, in 1941, and the Ph.D. degree from Oregon State College, Corvallis, in 1951. In 1951, he became a research assistant professor in electrical engineering at Syracuse University, Syracuse, N. Y., where he is now a senior research engineer. He has done research in high-voltage discharge phenomena, electromagnetic theory, and antenna studies.

M. K. Hu

Dr. Hu is a member of Sigma Xi, Pi Mu Epsilon, and the American Physical Society.

❖

Yueh-Ying Hu (S'48–A'51–SM'56) was born on May 11, 1918, in Nanking, China. She received the B.S. degree from National Central University, in Chungking, China, in 1940, and the Ph.D. degree from Oregon State College, Corvallis, in 1951.

Y. Y. Hu

From 1944 to 1947, she was an instructor at National Central University. In 1951, she became an assistant professor of electrical engineering at Syracuse University, Syracuse, N. Y., where she is now an associate professor.

Dr. Hu is a member of Sigma Xi, Pi Mu Epsilon, and the American Mathematical Society.

❖

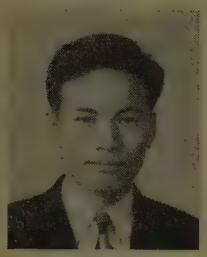
Richard F. Hyneman (S'49–A'54–M'58) was born on April 5, 1928, in Bloomington, Ind. He received the B.S. degree in 1950 from Purdue University, Lafayette, Ind., and the Ph.D. degree from the University of Illinois, Urbana, Ill., in February, 1957, both in electrical engineering.

R. F. Hyneman

From 1946 to 1948, he served in the U. S. Navy as an electronics technician. While at the University of Illinois, he received an appointment as research associate in the Antenna Laboratory, where he was engaged in studies of microwave antennas. From 1957 to April, 1959, he was employed by the Hughes Aircraft Company as a member of the technical staff of the Microwave Laboratory at Culver City, Calif. During this time, he was engaged in problems involving surface wave antennas and large radar slot arrays. He is presently head of the Microwave Analysis Section in the Radar Laboratory at Fullerton, Calif.

Dr. Hyneman is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.

Keigo Iizuka was born on August 29, 1931, in Kobe, Japan. He received the B.S. and M.S. degrees in electrical engineering, in 1955 and 1957, respectively, both from Kyoto University, Kyoto, Japan, He attended Harvard University, Cambridge, Mass., as a Fulbright exchange student from 1957 to 1958, receiving the M.S. degree in applied physics in 1958. He is now working toward the Ph.D. degree at Harvard University on a Gordon McKay Prize Scholarship.



K. Iizuka

Mr. Iizuka is a member of Sigma Xi and the Society of Electrical Communication Engineers of Japan.

❖

Wolfgang H. Kummer (S'46–A'52–SM'57) was born on October 10, 1925, in Stuttgart, Germany. He received the B.S. degree in 1946, the M.S. degree in 1947, and the Ph.D. degree in 1954, all in electrical engineering, from the University of California, Berkeley.

W. H. Kummer

From 1950 until 1953, he was lecturer in electrical engineering and a research engineer at the University of California Electronic Research Laboratory. In 1953, he joined the Radio Research Department of Bell Telephone Laboratories, Holmdel, N. J., where he performed experiments and analyses in radio wave propagation. Recently, he became associated with the Microwave Laboratory of Hughes Aircraft Company, Culver City, Calif., where he holds the position of staff engineer in the Antenna Research Department.

Dr. Kummer is a member of Eta Kappa Nu, Phi Beta Kappa, Tau Beta Pi, and Sigma Xi.

❖

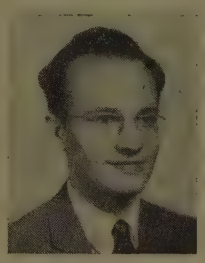
Paul A. Mann was born on September 13, 1912, in Pforzheim, Germany. He received the Ph.D. degree in theoretical physics in 1939 from the University of Munich, Germany.

P. A. Mann

From 1939 to 1945, he was employed by Telefunken Company for wireless telegraphy, Ltd, Berlin. After a short time in the patent office, he joined the television research group and was at first concerned with filter problems (Gaussian pulses) and then with theoretical investigations on noise problems. After World War II, he became an assistant at the Institute for Theoretical Physics of the University of Munich. Then, in 1951, he was again employed by Telefunken in Ulm, Germany. He is presently a member of the research group, and a lecturer, at the Technical High School of Stuttgart, Germany.

❖

Samuel P. Morgan (SM'55) was born in San Diego, Calif., on July 14, 1923. Ne received the B.S. degree in 1943, the M.S. degree in 1944, and the Ph.D. degree in physics in 1947, all from the California Institute of Technology, Pasadena.



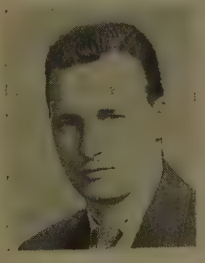
S. P. Morgan

Since 1947, he has been a research mathematician with Bell Telephone Laboratories, Murray Hill, N. J., working in various fields of mathematical physics. He has been particularly concerned with the applications of electromagnetic theory to microwave antennas and to problems of waveguide and coaxial cable transmission.

Dr. Morgan is a member of the American Physical Society, Tau Beta Pi, and Sigma Xi.

❖

Donald B. Muldrew was born in Winnipeg, Canada, on October 17, 1934. He received his B.Sc. degree in honors in Physics from the University of Manitoba, Winnipeg, in 1957.



D. B. Muldrew

Since that time, he has been at the Defence Research Telecommunitions Establishment, Ottawa, Canada, where he has worked on ionospheric problems, including determination of the spatial distribution of sporadic E and the calculation of focusing of radio waves by the ionosphere.

❖

Arthur A. Oliner (M'47–SM'52) was born in Shanghai, China, on March 5, 1921. He received the B.A. degree from Brooklyn College, N. Y., in 1941, and the Ph.D. degree in physics from Cornell University, Ithaca, N. Y., in 1946. While at Cornell, he held a graduate teaching assistantship in the physics department and also conducted research on an Office of Scientific Research and Development project.

A. A. Oliner

Since 1946, he has been with the Microwave Research Institute of the Polytechnic Institute of Brooklyn, N. Y., where he has been engaged in research in a variety of topics in the microwave field. He has also taught graduate courses in physics and electrical engineering, and is a research professor.

Dr. Oliner is Chairman of both the IRE Committee on Antennas and Waveguides and the IRE Professsional Group on Microwave and Theory and Techniques. He is a member of Commissions 1 and 6.3 of URSI, the American Physical Society, and Sigma Xi. He is also on the National Academy of Sciences Advisory Panel to the National Bureau of Standards.

❖

Hermann Poehler was born on September 10, 1928, in Ulm, Germany. He received the B.S. degree from the Technical University, Stuttgart, Germany, in 1951, and the M.S. degree in physics in 1958 from the Technical University in Munich, Germany, where he is presently studying for the Ph.D. degree.

H. Poehler

He was employed by the Air Force Cambridge Research Center, Bedford, Mass., from 1953 to 1957.

❖

S. R. Seshadri was born on October 25, 1928, in Madras, India. He graduated with the M.S. degree in physics in 1951 from the University of Madras, India, and he earned the diploma in electrical communication engineering in 1953 from the Indian Institute of Science, Bangalore, India. From 1954 to 1955 he served as a lecturer in electronics at the Madras Institute of Technology, Madras. He became a graduate student at Harvard University, Cambridge, Mass., in 1955, completing the work for the Ph.D. degree in applied physics in 1959. He is now a research fellow in electronics at Harvard University.

S. R. Seshadri

Dr. Seshadri is a member of Sigma Xi.

A. W. Straiton (M'47–SM'49–F'53) was born on August 27, 1907, in Arlington, Tex. He received the B.S.E.E. degree in 1929, the M.A. degree in 1931, and the Ph.D. degree in 1939, from the University of Texas, Austin.

A. W. STRAITON

He taught at Texas College of Arts and Industries from 1931 to 1943, before joining the staff of Electrical Engineering Department of the University of Texas, where he is present professor of electrical engineering and director of the Electrical Engineering Research Laboratory.

Dr. Straiton is a member of Commission II of URSI and is a former member of the U. S. National Committee.

❖

J. Irvin Swigart received the B.S. degree in physics at Illinois Wesleyan University, Bloomington, Ill., in 1929. He received the M.A. and the Ph.D. degrees in physics, in 1933 and 1938 respectively, both from Indiana State University.

J. I. SWIGART

From 1930 to 1931 he was a professor of physics at Bethany College, W. Va. In 1931 he became a physics instructor at the University of Utah, Salt Lake City, and since 1946 has been a full professor.

Dr. Swigart is a member of the American Association of University Professors, the Board of Directors of the American Rocket Society, the American Institute of Physics, Sigma Xi, the AIEE, the Utah Academy of Science, the American Association of Physics Teachers, Phi Eta Sigma, and Sigma Pi Sigma.

❖

Hugh T. Tomlinson was born in Waco, Tex., on April 22, 1925. He attended Baylor University, Waco, and the University of Texas, Austin, receiving the B.S. degree in electrical engineering in 1950, the M.S.E.E. degree in 1954, and the Ph.D. degree in 1950.

H. T. TOMLINSON

After twenty months' employment as a junior engineer with Westinghouse Electric Corporation, he returned to the University of Texas, in the Defense Research Laboratory, as a research engineer in the fields of microwave systems, propagation, and component development. From 1955 to 1957 he was a member of the electrical engineering staff at the University of Texas. He is presently employed by the Chance-Vought Aircraft Corporation, Dallas, Tex., as an electronics specialist in the field of component development.

Dr. Tomlinson is a member of Tau Beta Pi and Eta Kappa Nu.

INSTITUTIONAL LISTINGS

The IRE Professional Group on Antennas and Propagation is grateful for the assistance given by the firms listed below, and invites application for Institutional Listing from other firms interested in the field of Antennas and Propagation.

ANDREW CORPORATION, 363 E. 75th St., Chicago 19, Ill.
Antennas, Antenna Systems, Transmission Lines, Development and Production.

ANTLAB, INC., 6330 Proprietors Rd., Worthington, Ohio
Antenna Pattern Range Systems—Recorders & Mounts.

BLAINE ELECTRONETICS, INC., 14757 Keswick St., Van Nuys, Calif.
Antennas, Paraboloids, Scale Models, Antenna Radiation Pattern Measurement Towers.

DEVELOPMENTAL ENGINEERING CORP., 1001 Conn. Ave. N.W., Washington, D. C. and Leesburg, Va.
Research, Development, Installation of Antennas and Antenna Equipment for Super Power Stations.

THE GABRIEL LABORATORIES, Div. of the Gabriel Co., 135 Crescent Road, Needham Heights 94, Mass.
Research and Development of Antenna Equipment for Government and Industry.

HUGHES AIRCRAFT COMPANY, Culver City, Calif.
Research, Development, Mfr.: Radar, Missiles, Antennas, Radomes, Tubes, Solid State Physics, Computers.

I-T-E CIRCUIT BREAKER CO., Special Products Div., 601 E. Erie Ave., Philadelphia 34, Pa.
Design, Development and Manufacture of Antennas, and Related Equipment.

JANSKY & BAILEY, INC., 1339 Wisconsin Ave. N.W., Washington 7, D. C.
Radio & Electronic Engineering; Antenna Research & Propagation Measurements; Systems Design & Evaluation.

MARK PRODUCTS CO., 6412 W. Lincoln Ave., Morton Grove, Ill.
Multi Element Grid Parabolas, Antennas for Two-Way Communications, R & D.

THE W. L. MAXSON CORP., 475 Tenth Ave., New York 18, N.Y.
Research, Development, & Manufacture of Airborne, Missile & Ordnance Systems & Equipment.

TRANSCO PRODUCTS, INC., 12210 Nebraska Ave., Los Angeles 25, Calif.
Res., Design, Dev., & Mfr. of Antenna Systems & Components for Missile, Aircraft & Ground Installations.

WEINSCHEL ENGINEERING COMPANY, INC., Kensington, Md.
Antenna Pattern Receivers; Bolometer Amplifiers; Modulated Microwave Sources; Insertion Loss Measuring Systems

WHEELER LABORATORIES, INC., Great Neck, N. Y.; Antenna Lab., Smithtown, N. Y.
Consulting Services, Research and Development, Microwave Antennas and Waveguide Components.

WIND TURBINE COMPANY, West Chester, Pa.
Complete Antenna Systems and Towers

The charge for an Institutional Listing is $25.00 per issue or $75.00 for four consecutive issues. Application may be made to the Technical Secretary, The Institute of Radio Engineers, 1 East 79th Street, New York 21, N. Y.